

University of Groningen

## The neutral theory of biodiversity with random fission speciation

Etienne, Rampal S.; Haegeman, Bart

*Published in:*  
Theoretical ecology

*DOI:*  
[10.1007/s12080-010-0076-y](https://doi.org/10.1007/s12080-010-0076-y)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2011

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Etienne, R. S., & Haegeman, B. (2011). The neutral theory of biodiversity with random fission speciation. *Theoretical ecology*, 4(1), 87-109. <https://doi.org/10.1007/s12080-010-0076-y>

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

# The neutral theory of biodiversity with random fission speciation

Rampal S. Etienne · Bart Haegeman

Received: 9 October 2009 / Accepted: 10 March 2010 / Published online: 27 April 2010  
© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** The neutral theory of biodiversity and biogeography emphasizes the importance of dispersal and speciation to macro-ecological diversity patterns. While the influence of dispersal has been studied quite extensively, the effect of speciation has not received much attention, even though it was already claimed at an early stage of neutral theory development that the mode of speciation would leave a signature on metacommunity structure. Here, we derive analytical expressions for the distribution of abundances according to the neutral model with recruitment (i.e., dispersal and establishment) limitation and random fission speciation which seems to be a more realistic description of (allopatric) speciation than the point mutation mode of speciation mostly used in neutral models. We find that the two modes of speciation behave qualitatively differently except when recruitment is strongly limited. Fitting the model to six large tropical tree data sets, we show that it performs worse than the original neutral model with point mutation speciation but yields more realistic predictions for speciation rates, species longevities, and rare species. Interestingly, we find that the metacommunity abundance distribution under random fission is identical to the broken-stick abundance distribution

and thus provides a dynamical explanation for this grand old lady of abundance distributions.

**Keywords** Sampling formula · Maximum likelihood · Fundamental biodiversity number

## Introduction

Understanding the assembly and biodiversity of ecological communities is the primary aim of community ecology. In an excellent minireview of ecological assembly rules, Belyea and Lancaster (1999) summarized the main drivers or constraints determining community structure: environmental constraints, dispersal constraints, internal dynamics (such as competition), and biogeography (which includes the processes of speciation and extinction). Most of the literature in the past decades has focused on the first three factors (see, e.g., Cody and Diamond 1975; Weiher and Keddy 2001). In contrast, biogeography and the processes of diversification have received relatively little attention despite the pioneering work by MacArthur and Wilson (1967), Collwell and Winkler (1984), and Ricklefs (1987) who clearly showed the importance of these processes for community composition and macro-ecological patterns. However, there has been a revived interest in the influence of biogeographical forces on community structure in the last decade due to two new developments. The first development is community phylogenetics, which studies evolutionary relationships between coexisting species (Webb et al. 2002; Cavender-Bares et al. 2009). The second is neutral community ecology, which suspends the role of species differences in order to create a null model that allows

R. S. Etienne (✉)  
Community and Conservation Ecology Group,  
Centre for Ecological and Evolutionary Studies,  
University of Groningen, P.O. Box 14,  
9750 AA Haren, The Netherlands  
e-mail: r.s.etienne@rug.nl

B. Haegeman  
INRIA Research Team MERE,  
UMR Systems Analysis and Biometrics,  
2 place Pierre Viala, 34060 Montpellier, France

the study of factors other than asymmetrical species interactions such as dispersal and biogeography (Hubbell 2001). In this paper, we focus on the latter development and particularly on the impact of speciation on the shape of species abundance distributions.

Building on the classic works of MacArthur and Wilson (1967) and Caswell (1976), Hubbell (2001) introduced his neutral theory of community ecology that states that stochastic interplay between a few basic, ecological as well as macro-evolutionary, processes (speciation, birth, and death, and—on a local scale—dispersal) can explain general large-scale diversity patterns, such as species-abundance distributions (SADs) and species–area curves. The neutral theory as developed by Hubbell (2001) makes three basic assumptions: (1) individuals of different species are functionally equivalent (neutrality assumption), (2) the community size is constant (zero-sum assumption), and (3) speciation is comparable to mutation where each individual has an equal probability of producing mutated, i.e., speciated, offspring (point mutation assumption). While the neutrality assumption is at the heart of the theory, the other two assumptions are assumptions of particular model implementations of the theory, allowing for analytical expressions for diversity measures, rather than fundamental assumptions of the theory itself. A mismatch between observations and theoretical predictions can, in principle, be due to these additional assumptions and therefore cannot be immediately interpreted as calling for a rejection of the neutral theory as a whole (Etienne 2007).

The zero-sum assumption and the point mutation assumption are of a very different nature. It has been shown that models *without* the zero-sum assumption predict mathematically exactly the same equilibrium SAD as the model *with* this assumption (Etienne et al. 2007a; Haegeman and Etienne 2008; Conlisk et al. 2010). In contrast, alternatives to the point mutation assumption can predict very different SADs (Hubbell 2001; Etienne et al. 2007b; Allen and Savage 2007; Haegeman and Etienne 2009; De Aguiar et al. 2009). Indeed, Hubbell (2001) claimed that speciation would leave a signature on diversity patterns in the metacommunity (see also Mouillot and Gaston 2007). Therefore, a thorough analysis of neutral theory, or any theory of community assembly for that matter, requires the (quantitative) exploration of alternative modes of speciation, particularly those that are very different from point mutation. Hubbell (2001) proposed an alternative speciation mode that is the opposite of the sympatric point mutation mode which he dubbed “random fission” because speciation results from random splitting of populations which may be interpreted as

mimicking allopatric speciation. Intuitively, it seems more reasonable than point mutation because the incipient abundance of new species is larger than a single individual (Allen and Savage 2007; Rosindell et al. 2010), and it is also more plausible than a fixed initial abundance, as assumed by Allen and Savage (2007). Hubbell (2001) stated that the equilibrium *metacommunity* SAD resulting from random fission speciation is a zero-sum multinomial, just like the *local* community SAD in the point mutation case—which was later called dispersal-limited (Etienne and Alonso 2005, 2007), or recruitment-limited (Jabot et al. 2008) multinomial—but he did not prove this mathematically. Ricklefs (2003) provided some approximate formulas for the total species richness in the metacommunity under random fission speciation, but so far, a full mathematical treatment has remained elusive.

In this paper, we provide the full sampling formula for the distribution of abundance in a local community that receives immigrants from a very large metacommunity described by random fission speciation. It is thus the counterpart of the sampling formula where the metacommunity is described by point mutation (Etienne 2005) and may be similarly extended to involve multiple samples (Etienne 2007, 2009a, b). The metacommunity SAD is clearly different from the zero-sum multinomial, in contrast to Hubbell’s conjecture. We use the new sampling formula to fit the random fission model to six large tropical tree data sets and compare it to the fit of the point mutation model. We end with a discussion of our results.

## Model

We will first describe metacommunity dynamics, solve it for the stationary abundance distribution, and then derive expressions for (possibly dispersal-limited) local samples from this stationary distribution. We add the superscript “meta” to expectations and probabilities that refer to the metacommunity to distinguish them from expressions for samples, which we will denote by the superscript “smp”.

### The master equation

We follow Hubbell (2001) in assuming a constant metacommunity size and denote it by  $J_M$ . However, as in the point mutation model (Etienne et al. 2007a; Haegeman and Etienne 2008), this assumption is not essential for determining the equilibrium species abundance distribution because fluctuations of species abundances cancel out and yield an effectively constant

community size, that is, a sharply peaked probability distribution for metacommunity size (Haegeman and Etienne 2010). We denote the abundance vector in the metacommunity by  $\vec{S} = (S_1, S_2, S_3, \dots, S_{J_M})$  where each component  $S_n$  is the number of species with abundance  $n$ . We study the behavior of the probability  $\mathbb{P}^{\text{meta}}(\vec{S})$  of this vector in time by writing down the so-called master equation for this vector  $\vec{S}$  which reads, in general (Haegeman and Etienne 2009):

$$\frac{d}{dt} \mathbb{P}^{\text{meta}}(\vec{S}) = \sum_{\vec{S}' \neq \vec{S}} \left( \mathbb{P}^{\text{meta}}(\vec{S}') R(\vec{S}', \vec{S}) - \mathbb{P}^{\text{meta}}(\vec{S}) R(\vec{S}, \vec{S}') \right) \quad (1)$$

where  $R$  is a matrix that contains the rates of transitions  $R(\vec{S}, \vec{S}')$  to go from a vector  $\vec{S}$  to another vector  $\vec{S}'$ . The master equation (Eq. 1) describes the dynamics of the species abundance distribution  $\mathbb{P}^{\text{meta}}(\vec{S})$  as a function of time. This distribution contains much more information than just the expectation values  $\mathbb{E}^{\text{meta}}(S_n)$ , used by, for example, Vallade and Houchmandzadeh (2003) and Volkov et al. (2003, 2007), which is just the first moment of this distribution (Vanpeteghem et al. 2008), as we will see below.

To specify the stochastic community model, one has to determine the matrix  $R$  of transition rates. For different models, the matrix  $R$  will have different elements. For the random fission model which seems impossible, there are two types of transitions, and the matrix  $R$  can be written as a sum of two parts:  $R = R^{\text{DB}} + R^{\text{S}}$ . The first part,  $R^{\text{DB}}$ , describes a death event immediately followed by a birth event. In such a death–birth event, one individual dies (all  $J_M$  individuals have the same probability to be the one that dies), and a second individual reproduces (all remaining  $J_M - 1$  individuals have the same probability to reproduce). Note that this is exactly the death–birth event of the point mutation model. If the first individual belongs to a species with abundance  $k$  and the second individual to a species with abundance  $i$ , the transition is to a vector  $\vec{S}'$  that can be written in terms of vector  $\vec{S}$  as

$$\vec{S}' = \vec{S}_{k,i}^{\text{DB}} = \vec{S} - \vec{e}_k + \vec{e}_{k-1} - \vec{e}_i + \vec{e}_{i+1} \quad (2)$$

which means that it differs from vector  $\vec{S}$  by having a reduction by one of species with abundances  $k$  and  $i$  and an increase by one of species with abundances  $k - 1$  and  $i + 1$  (the unit vector  $\vec{e}_k$  is one at position  $k$  and zero elsewhere). The corresponding element of the

transition matrix  $R^{\text{DB}}(\vec{S}, \vec{S}_{k,i}^{\text{DB}})$ , for which we use the shorthand notation  $R_{k,i}^{\text{DB}}(\vec{S})$ , is given by

$$R^{\text{DB}}(\vec{S}, \vec{S}_{k,i}^{\text{DB}}) = R_{k,i}^{\text{DB}}(\vec{S}) = \mu \frac{k S_k}{J_M} \frac{i S_i}{J_M - 1} \quad (3)$$

with  $\mu$  the overall rate of death–birth events. Strictly speaking, the transition rate 3 is only valid for  $k \neq i$  (see “Appendix 1” for the full derivation).

The second part of matrix  $R$ , denoted by  $R^{\text{S}}$ , describes a speciation event. We describe a speciation event as in Hubbell’s (2001) random fission model: An individual is selected at random (all  $J_M$  individuals have the same probability to be selected). Suppose that this individual belongs to a species with abundance  $k$ . Then this species splits into two species, the first with abundance  $i$  (where  $i < k$ ) and the second with abundance  $k - i$ , where each fission has equal probability, i.e., each split  $(i, k - i)$  has probability  $\frac{1}{k-1}$ . The transition is to a vector  $\vec{S}'$  that can be written in terms of vector  $\vec{S}$  as

$$\vec{S}' = \vec{S}_{k,i}^{\text{S}} = \vec{S} - \vec{e}_k + \vec{e}_i + \vec{e}_{k-i} \quad (4)$$

which means that it differs from vector  $\vec{S}$  by having one less species of abundance  $k$  and an increase in the number of species with abundances  $i$  and  $k - i$ . The corresponding element of the transition matrix  $R^{\text{S}}(\vec{S}, \vec{S}_{k,i}^{\text{S}})$ , for which we use the shorthand notation  $R_{k,i}^{\text{S}}(\vec{S})$ , is given by

$$R_{\text{rf}}^{\text{S}}(\vec{S}, \vec{S}_{k,i}^{\text{S}}) = R_{k,i}^{\text{S}}(\vec{S}) = \nu \frac{k S_k}{J_M} \frac{1}{k-1} \quad (5)$$

with  $\nu$  the overall speciation rate.

For comparison, in the point mutation model, the death–birth events are also governed by Eq. 3, but speciation only produces singleton species, that is, only transitions to  $i = 1$  are allowed:

$$R_{\text{pm}}^{\text{S}}(\vec{S}, \vec{S}_{k,i}^{\text{S}}) = R_{k,i}^{\text{S}}(\vec{S}) = \nu \frac{k S_k}{J_M} \delta_1(i) \quad (6)$$

where  $\delta_1(i) = 1$  for  $i = 1$  and 0 otherwise.

The master equation (Eq. 1) together with the transition rates 3 and 5 fully define the neutral model with random fission speciation. Rather than trying to solve Eq. 1 for the random fission model which seems impossible, we take an indirect route consisting of three steps. First, we derive an equation for the expected number of species with abundance  $n$  which we will denote by  $\mathbb{E}^{\text{meta}}(S_n)$ . We can solve this equation exactly. This will illustrate the main properties of the random fission model, but cannot be used to fit the model to data. To the latter end, we need the full sampling formula for the abundance distribution of a local community connected via limited dispersal to a metacommunity

governed by random fission speciation. The second step starts by proposing an Ansatz for the solution of Eq. 1 and show that it is consistent with the previously obtained exact expression for  $\mathbb{E}^{\text{meta}}(S_n)$ . The third and final step applies sampling theory (Etienne and Alonso 2005) to this Ansatz, assuming large metacommunity size, in order to formulate the full sampling formula.

Equation for the expected number of species with abundance  $n$

The definition of the expected number of species with abundance  $n$  in the metacommunity,  $\mathbb{E}^{\text{meta}}(S_n)$ , is

$$\mathbb{E}^{\text{meta}}(S_n) = \sum_{\vec{S}} S_n \mathbb{P}^{\text{meta}}(\vec{S}) \quad (7)$$

It is straightforward (see “Appendix 1”), by taking time derivatives on both sides of this equation and using Eq. 1 to find the stationary expectation of the number of species with abundance  $n$  in the metacommunity:

$$\mathbb{E}^{\text{meta}}(S_n) = \prod_{i=1}^n a_i \quad (8a)$$

with

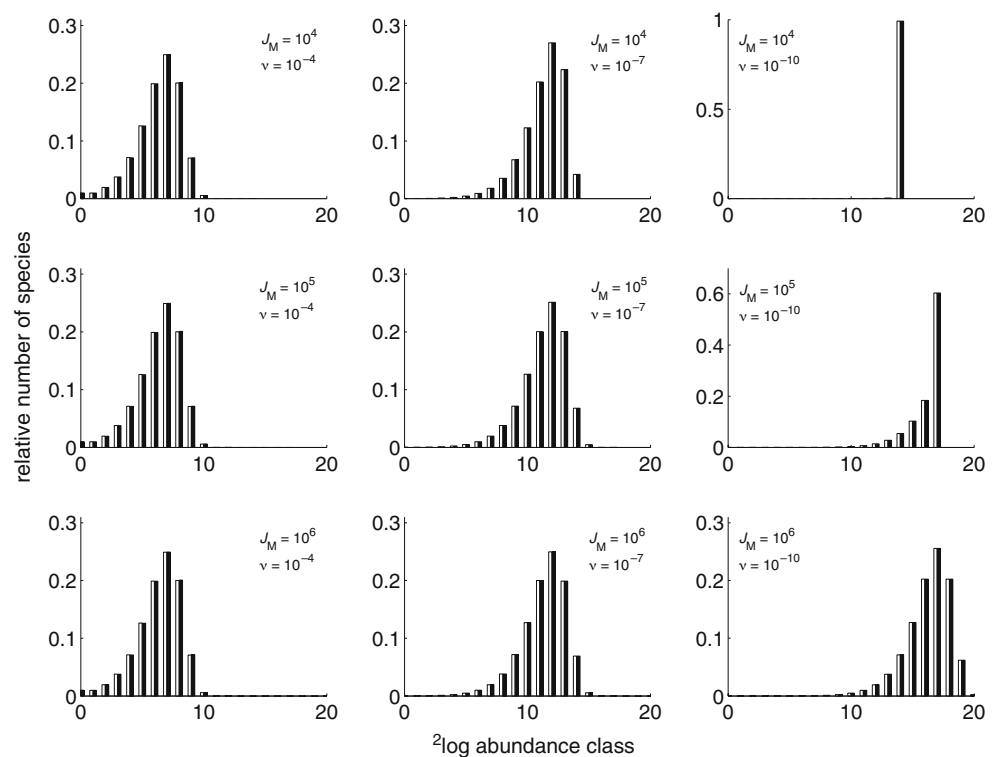
$$\begin{cases} a_1 = \frac{\nu}{r_1 + s_1} \\ a_i = \frac{r_{i-1}}{2r_i + s_i - r_{i+1}a_{i+1} - \sum_{j=i+1}^{J_M} s_j \frac{2}{j-1} \prod_{k=i+1}^j a_k} \\ \text{for } 1 < i < J_M \\ a_{J_M} = \frac{r_{J_M-1}}{r_{J_M} + s_{J_M}} \end{cases} \quad (8b)$$

It is easy to see that Eq. 8 can be computed numerically by first calculating all  $a_i$  for  $i$  from  $J_M$  down to 1 and then by calculating  $\mathbb{E}^{\text{meta}}(S_n)$  for  $n$  from 1 to  $J_M$ .

Note that these expressions only depend on  $J_M$  and the ratio of  $\nu$  and  $\mu$ , that is,  $\frac{\nu}{\mu}$  which is the relative rate of speciation and birth–death events, similar to the point mutation model (Vallade and Houchmandzadeh 2003; Etienne et al. 2007a). Therefore, we will assume for simplicity of notation that  $\mu = 1$  without loss of generality. Figure 1 shows some numerical examples of  $\mathbb{E}^{\text{meta}}(S_n)$  for various values of  $\nu$  and  $J_M$ .

So far, we have been able to provide exact solutions for the expectation values of  $S_n$  for the stationary abundance distribution. Although these can be used for comparison to data after incorporating sampling, it does not allow full extraction of the information in

**Fig. 1** Comparison of exact (Eq. 8, white) and approximate (Eq. 17a, black) predictions for the relative number of species in each abundance class, i.e.,  $\mathbb{E}^{\text{meta}}(\sum_{n \in \text{class } i} S_n) / \mathbb{E}^{\text{meta}}(S)$ . Abundance classes are defined on a logarithmic scale: the  $i$ th class being  $[2^{i-1}, 2^i]$ , e.g., the first class contains abundance 1, the second class abundances 2 and 3, the third abundances 4, 5, 6, and 7, the fourth abundances 8 through 15, etc. (Pueyo 2006)



a sample abundance vector  $\vec{S}$ . For this, the full probability distribution, rather than just the first moment, is required. This full probability, called the sampling formula, enables one to extract information on individual species' abundances as well as on their interdependencies. We therefore proceed to derive an extremely good approximation to this sampling formula. We can check the accuracy of this approximation by comparing the expectation value  $\mathbb{E}^{\text{meta}}(S_n)$  computed using this approximation (and higher moments) to the exact solution for  $\mathbb{E}^{\text{meta}}(S_n)$ .

Ansatz for the probability of the metacommunity abundance vector  $\vec{S}$

We start with an Ansatz which is justified elsewhere (Haegeman and Etienne 2010). Here, we only note that it is based on an approximation, for large  $J_M$ , of the exact formulas 8 and below we will show numerically that it is an extremely good approximation for realistic values of  $J_M$  and  $\nu$  which justifies our use of it. The Ansatz is the following expression for the probability of the metacommunity abundance distribution  $\vec{S}$  given a fixed size  $J_M$ :

$$\begin{aligned}\mathbb{P}^{\text{meta}}(\vec{S}|J_M) &= \frac{1}{Z(\nu J_M, J_M)} \prod_{k=1}^{J_M} \frac{(\nu J_M)^{S_k}}{S_k!} \\ &= \frac{(\nu J_M)^{S_M}}{Z(\nu J_M, J_M)} \prod_{k=1}^{J_M} \frac{1}{S_k!}\end{aligned}\quad (9a)$$

where  $S_M$  is the total number of species in the metacommunity and  $S_k$  is the number of species with abundance  $k$ , both being stochastic variables as above (in contrast to  $\mathbb{E}^{\text{meta}}(S_n)$  and  $\mathbb{E}^{\text{meta}}(S_M)$  which are moments). The normalization constant  $Z(\nu J_M, J_M)$  is given by

$$\begin{aligned}Z(\nu J_M, J_M) &= \sum_{\vec{S}|J_M} \prod_{k=1}^{J_M} \frac{(\nu J_M)^{S_k}}{S_k!} \\ &= \sum_{S_M=1}^{J_M} \frac{(\nu J_M)^{S_M}}{S_M!} \sum_{\vec{S}|J_M, S_M} \frac{S_M!}{\prod_{k=1}^{J_M} S_k!} \\ &= \sum_{S_M=1}^{J_M} \frac{(\nu J_M)^{S_M}}{S_M!} \binom{J_M-1}{S_M-1} \\ &= \nu J_M {}_1F_1(1 - J_M, 2, -\nu J_M)\end{aligned}\quad (9b)$$

where  ${}_1F_1(a, b, x)$  is the confluent hypergeometric function and  $\binom{x}{y} = \frac{x!}{y!(x-y)!}$  is the usual notation of the

binomial coefficient. The second line is due to a classic combinatorial result,

$$\sum_{\vec{S}|J_M, S_M} \binom{S_M}{\vec{S}} = \binom{J_M-1}{S_M-1} \quad (10)$$

From Eq. 9, we can derive (see “Appendix 2”) that

$$\mathbb{E}^{\text{meta}}(S_n|J_M) = \nu J_M \frac{Z(\nu J_M, J_M - n)}{Z(\nu J_M, J_M)} \quad (11)$$

and that

$$\begin{aligned}\mathbb{P}^{\text{meta}}(S_M|J_M) &= \sum_{\vec{S}|J_M, S_M} \mathbb{P}^{\text{meta}}(\vec{S}|J_M) \\ &= \frac{1}{Z(\nu J_M, J_M)} \frac{(\nu J_M)^{S_M}}{S_M!} \binom{J_M-1}{S_M-1}\end{aligned}\quad (12)$$

Hence

$$\begin{aligned}\mathbb{E}^{\text{meta}}(S_M|J_M) &= \sum_{S_M=1}^{J_M} S_M \mathbb{P}^{\text{meta}}(S_M|J_M) \\ &= \sum_{S_M=1}^{J_M} S_M \frac{(\nu J_M)^{S_M}}{S_M!} \binom{J_M-1}{S_M-1} \\ &= \frac{{}_1F_1(1 - J_M, 1, -\nu J_M)}{{}_1F_1(1 - J_M, 2, -\nu J_M)}\end{aligned}\quad (13)$$

Furthermore, we find that

$$\begin{aligned}\mathbb{P}^{\text{meta}}(\vec{S}|J_M, S_M) &= \frac{\mathbb{P}^{\text{meta}}(\vec{S}|J_M)}{\mathbb{P}^{\text{meta}}(S_M|J_M)} \\ &= \frac{1}{\binom{J_M-1}{S_M-1}} \frac{S_M!}{\prod_k S_k!}\end{aligned}\quad (14)$$

This is a very interesting result because it corresponds exactly to the discrete version of MacArthur's (1957) broken-stick model of the distribution of species abundances (see Etienne and Olff 2005)!

Scaling limit

We consider the scaling  $J_M \rightarrow \infty$ ,  $\nu \rightarrow 0$  such that  $\sqrt{\nu} J_M$  is finite. We call the quantity  $\sqrt{\nu} J_M$  the fundamental biodiversity number in the random fission model and denote it by  $\theta_{\text{rf}}$ . We first study the distrib-



ution of the number of species  $S_M$ . In the scaling limit, Eq. 12 becomes

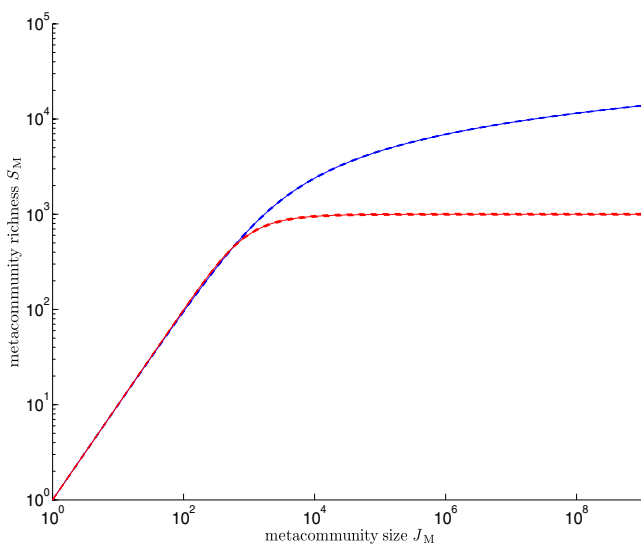
$$\mathbb{P}^{\text{meta}}(S_M | J_M \rightarrow \infty) = \frac{1}{Z(\theta_{\text{rf}})} \frac{\theta_{\text{rf}}^{2S_M}}{S_M! (S_M - 1)!} \quad (15a)$$

with normalization constant  $Z(\theta_{\text{rf}})$  given by

$$Z(\theta_{\text{rf}}) = \sum_{S_M=1}^{\infty} \frac{\theta_{\text{rf}}^{2S_M}}{S_M! (S_M - 1)!} = \theta_{\text{rf}} I_1(2\theta_{\text{rf}}) \quad (15b)$$

where  $I_\alpha(x)$  denotes the modified Bessel function of the first kind for integer  $\alpha$  and real-valued  $x$ ; this is a standard mathematical function for which most mathematical software packages have numerical routines. Figure 2 shows that the mean and the mean  $\pm 1$  standard deviation of this distribution for  $S_M$  as a function of  $J_M$  almost coincide, indicating that the distribution is strongly peaked. This result can be obtained analytically; we refer to “Appendix 3” for the proof that, for large  $J_M$  and large  $\theta_{\text{rf}}$ , the number of species  $S_M$  is normally distributed with mean and variance equal to  $\theta_{\text{rf}}$  and  $\frac{\theta_{\text{rf}}}{2}$ , respectively. For any  $\theta_{\text{rf}}$  (but still in the limit  $J_M \rightarrow \infty$ ), we find that the expected number of species in the metacommunity is given by (see “Appendix 3”):

$$\mathbb{E}^{\text{meta}}(S_M | J_M \rightarrow \infty) = \theta_{\text{rf}} \frac{I_0(2\theta_{\text{rf}})}{I_1(2\theta_{\text{rf}})} \quad (16)$$



**Fig. 2** The expected total metacommunity richness as a function of metacommunity size for the random fission model (red) and the point mutation model (blue). Dashed lines indicate the variation in richness around the mean ( $\pm 1$  standard deviation). Mean and standard deviations are computed from the full probability distribution, Eqs. 29 and 30. The fundamental biodiversity constant is set at  $\theta_{\text{pm}} = \theta_{\text{rf}} = 1,000$

The expectation values of  $S_n, S_k S_\ell, \dots$  for large  $J_M$  can then be derived (see again “Appendix 4” for the full derivation)

$$\mathbb{E}^{\text{meta}}(S_n | J_M) \approx \frac{\theta_{\text{rf}}^2}{J_M \sqrt{1 - \frac{n}{J_M}}} \frac{I_1(2\theta_{\text{rf}} \sqrt{1 - \frac{n}{J_M}})}{I_1(2\theta_{\text{rf}})} \quad (17a)$$

for large  $J_M$

$$\mathbb{E}^{\text{meta}}(S_k S_\ell | J_M) \approx \frac{\theta_{\text{rf}}^4}{J_M^2 \sqrt{1 - \frac{k+\ell}{J_M}}} \frac{I_1(2\theta_{\text{rf}} \sqrt{1 - \frac{k+\ell}{J_M}})}{I_1(2\theta_{\text{rf}})} \quad (17b)$$

for large  $J_M$

These approximations match exact numerical results for the random fission model outlined above very well. For example, Fig. 1 shows the extremely good match of the approximate expectations  $\mathbb{E}^{\text{meta}}(S_n)$  of Eq. 17a with the exact expectation given in Eq. 8. Similarly, the approximate second-order moments  $\mathbb{E}^{\text{meta}}(S_n S_m)$  of Eq. 17b are very close to the exact second-order moments  $\mathbb{E}^{\text{meta}}(S_n S_m)$ , solutions of Eq. 1 (results not shown). Because of this extremely good agreement, we believe that we can use the approximations with great confidence to explore the stationary properties of the random fission model.

For the derivation of our sampling formula for  $\mathbb{P}^{\text{smp}}(\vec{S} | I, \theta_{\text{rf}}, J)$  below, we need the probability distribution for the relative abundance vector  $\vec{p}$ . Equation 17 for large  $J_M$  can be transformed in probability density functions  $\rho^{\text{meta}}(p_i | S_M), \rho^{\text{meta}}(p_i, p_j | S_M), \dots$  for relative abundances  $p_i, p_j, \dots$  in the limit  $J_M \rightarrow \infty$  as follows: The probability that an individual belongs to a species having abundance  $p_i J_M$  is  $\frac{\mathbb{E}^{\text{meta}}(S_{p_i J_M} | J_M, S_M)}{S_M}$ . This probability becomes  $\rho^{\text{meta}}(p_i | S_M) dp_i$  in the limit  $J_M \rightarrow \infty$  (with  $dp_i \sim \frac{1}{J_M}$ ). Therefore,

$$\begin{aligned} \rho^{\text{meta}}(p_i | S_M) &= \lim_{J_M \rightarrow \infty} J_M \frac{\mathbb{E}^{\text{meta}}(S_{p_i J_M} | J_M, S_M)}{S_M} \\ &= (S_M - 1)(1 - p_i)^{S_M - 2} \end{aligned} \quad (18a)$$

Likewise,

$$\begin{aligned} \rho^{\text{meta}}(p_i, p_j | S_M) &= \lim_{J_M \rightarrow \infty} J_M^2 \frac{\mathbb{E}^{\text{meta}}(S_{p_i J_M} S_{p_j J_M} | J_M, S_M)}{S_M(S_M - 1)} \\ &= (S_M - 1)(S_M - 2)(1 - p_i - p_j)^{S_M - 3} \end{aligned} \quad (18b)$$

Continuing this procedure, we find that

$$\rho^{\text{meta}}(\vec{p} | S_M) = (S_M - 1)! \quad (18c)$$

Hence, all relative abundance vectors  $\vec{p}$  on  $S_M$  species are equally probable. Note that densities 18a and 18b can be obtained from Eq. 18c by computing marginal distributions.

The results for the metacommunity are summarized in Table 1 where they can be compared with the results for the point mutation model.

Sampling formula  $\mathbb{P}^{\text{smp}}(\vec{S})$  and sample expectations  $\mathbb{E}^{\text{smp}}(S_n)$

The asymptotic formula 18c can be used to derive our central result, i.e., the dispersal-limited sampling for-

mula (Alonso and McKane 2004; Etienne and Alonso 2005) for species abundances in a sample of size  $J$  from a dispersal-limited local community (parametrized by parameter  $I$ ) that receives immigrants from a metacommunity undergoing random fission speciation (parametrized by parameter  $\theta_{\text{rf}}$ ). The sampling formula gives the probability for the vector  $\vec{S} = (S_1, S_2, \dots, S_J)$ , where each component  $S_n$  denotes the number of species with abundance  $n$ , but it will be convenient to also use in our notation the abundance vector  $\vec{n} = (n_1, n_2, \dots, n_S)$  where each component  $n_i$  denotes the number of individuals of species  $i$  where the species are arranged in an arbitrary order. The detailed derivation of the sampling formula can be found in “Appendix 5”.

**Table 1** Comparison of random fission and point mutation formulas for the metacommunity

Quantity	Formula	
$\theta_{\text{rf}}$	$\sqrt{v_{\text{rf}}} J_M$	(19)
$\theta_{\text{pm}}$	$v_{\text{pm}} (J_M - 1) \approx v_{\text{pm}} J_M$	(20)
$\mathbb{P}_{\text{rf}}^{\text{meta}}(\vec{S} J_M)$	$\frac{1}{Z(v_{\text{rf}}J_M, J_M)} \prod_k \frac{(v_{\text{rf}}J_M)^{S_k}}{S_k!} = \frac{1}{Z(\frac{\theta_{\text{rf}}^2}{J_M}, J_M)} \prod_k \frac{\theta_{\text{rf}}^{2S_k}}{J_M^{S_k} S_k!}$	(21)
$\mathbb{P}_{\text{pm}}^{\text{meta}}(\vec{S} J_M)$	$\frac{J_M!}{(v_{\text{pm}}J_M)_{J_M}} \prod_k \frac{(v_{\text{pm}}J_M)^{S_k}}{k^{S_k} S_k!} = \frac{J_M!}{(\theta_{\text{pm}})_{J_M}} \prod_k \frac{\theta_{\text{pm}}^{S_k}}{k^{S_k} S_k!}$	(22)
$\mathbb{E}_{\text{rf}}^{\text{meta}}(S_n J_M)$	$v_{\text{rf}}J_M \frac{Z(v_{\text{rf}}J_M, J_M - n)}{Z(v_{\text{rf}}J_M, J_M)} \approx \frac{\theta_{\text{rf}}^2}{J_M \sqrt{1 - \frac{n}{J_M}}} \frac{I_1\left(2\theta_{\text{rf}}\sqrt{1 - \frac{n}{J_M}}\right)}{I_1(2\theta_{\text{rf}})}$	(23)
$\mathbb{E}_{\text{pm}}^{\text{meta}}(S_n J_M)$	$\binom{J_M}{n} \theta_{\text{pm}} \frac{(n-1)! (\theta_{\text{pm}})_{J_M-n}}{(\theta_{\text{pm}})_{J_M}} \approx \frac{\theta_{\text{pm}}}{n} \left(1 - \frac{n}{J_M}\right)^{\theta_{\text{pm}}-1}$	(24)
$\mathbb{P}_{\text{rf}}^{\text{meta}}(\vec{S} J_M, S_M)$	$\frac{S_M!}{(J_M-1)_{(S_M-1)}} \prod_k \frac{1}{S_k!}$	(25)
$\mathbb{P}_{\text{pm}}^{\text{meta}}(\vec{S} J_M, S_M)$	$\frac{J_M!}{\bar{s}(J_M, S_M)} \prod_k \frac{1}{k^{S_k} S_k!}$	(26)
$\mathbb{E}_{\text{rf}}^{\text{meta}}(S_n J_M, S_M)$	$\frac{S_M}{(J_M-1)_{(S_M-1)}} \binom{J_M-n-1}{S_M-2} \approx \frac{S_M(S_M-1)}{J_M} \left(1 - \frac{n}{J_M}\right)^{S_M-2}$	(27)
$\mathbb{E}_{\text{pm}}^{\text{meta}}(S_n J_M, S_M)$	$\binom{J_M}{n} \frac{(n-1)! \bar{s}(J_M-n, S_M-1)}{\bar{s}(J_M, S_M)}$	(28)
$\mathbb{P}_{\text{rf}}^{\text{meta}}(S_M J_M)$	$\frac{1}{Z(v_{\text{rf}}J_M, J_M)} \frac{(v_{\text{rf}}J_M)^{S_M}}{S_M!} \binom{J_M-1}{S_M-1} \approx \frac{1}{Z(\theta_{\text{rf}})} \frac{\theta_{\text{rf}}^{2S_M}}{S_M! (S_M-1)!}$	(29)
$\mathbb{P}_{\text{pm}}^{\text{meta}}(S_M J_M)$	$\bar{s}(J_M, S_M) \frac{\theta_{\text{pm}}^{S_M}}{(\theta_{\text{pm}})_{J_M}}$	(30)
$\mathbb{E}_{\text{rf}}^{\text{meta}}(S_M J_M)$	$\frac{{}_1F_1(1 - J_M, 1, -v_{\text{rf}}J_M)}{{}_1F_1(1 - J_M, 2, -v_{\text{rf}}J_M)} \approx \theta_{\text{rf}} \frac{I_0(2\theta_{\text{rf}})}{I_1(2\theta_{\text{rf}})}$	(31)
$\mathbb{E}_{\text{pm}}^{\text{meta}}(S_M J_M)$	$\sum_{k=1}^{J_M} \frac{\theta_{\text{pm}}}{\theta_{\text{pm}} + k - 1} = \theta_{\text{pm}} (\Psi(\theta_{\text{pm}} + J_{\text{pm}}) - \Psi(\theta_{\text{pm}}))$	(32)



Here, we mention the result (see also Eq. 42 in Table 2):

$$\mathbb{P}^{\text{sm}}(\vec{S}|I, \theta_{\text{rf}}, J) = \binom{J}{\vec{n}} \frac{1}{(I)_J \prod_k S_k!} \sum_{a_1=1}^{n_1} \cdots \sum_{a_S=1}^{n_S} \left[ \prod_{i=1}^S \bar{s}(n_i, a_i) a_i! \right] \times I^A \theta_{\text{rf}}^{S-A} \frac{I_{A+S-1}(2\theta_{\text{rf}})}{I_1(2\theta_{\text{rf}})} \quad (33)$$

Here, the  $I$  with a subscript denote modified Bessel functions, and the  $I$  without a subscript denotes the dispersal limitation parameter. Furthermore,  $A = \sum_{i=1}^S a_i$ ,  $a_i$  just being an index for the summations in Eq. 33,

$S$  is the number of species in the sample, and  $S_k$  is the number of species in the dataset that have abundance  $k$ . The  $\bar{s}(x, y)$  are the unsigned Stirling numbers of the first kind and  $(x)_y$  is the Pochhammer symbol defined as

$$(x)_y = \prod_{i=1}^y (x+i-1) = \frac{\Gamma(x+y)}{\Gamma(x)} = \sum_{i=1}^y \bar{s}(y, i) x^i \quad (34)$$

and the multinomial coefficient  $\binom{J}{\vec{n}}$  is defined as

$$\binom{J}{\vec{n}} = \frac{J!}{\prod_i n_i!} \quad (35)$$

**Table 2** Comparison of random fission and point mutation formulas for a sample of size  $J$

Quantity	Formula	
$\theta_{\text{rf}}$	$\sqrt{v_{\text{rf}}} J_{\text{M}}$	(36)
$\theta_{\text{pm}}$	$v_{\text{pm}}(J_{\text{M}} - 1) \approx v_{\text{pm}} J_{\text{M}}$	(37)
Metacommunity	(No dispersal limitation)	
$\mathbb{P}_{\text{rf}}^{\text{sm}}(\vec{S} \theta_{\text{rf}}, J)$	$\frac{J!}{\prod_k S_k!} \theta_{\text{rf}}^{S-J} \frac{I_{J+S-1}(2\theta_{\text{rf}})}{I_1(2\theta_{\text{rf}})}$	(38)
$\mathbb{P}_{\text{pm}}^{\text{sm}}(\vec{S} \theta_{\text{pm}}, J)$	$\frac{J!}{\prod_i n_i \prod_k S_k!} \frac{\theta_{\text{pm}}^S}{(\theta_{\text{pm}})_J}$	(39)
$\mathbb{P}_{\text{rf}}^{\text{sm}}(S_n \theta_{\text{rf}}, J)$	$\begin{cases} \frac{J!}{I_1(2\theta_{\text{rf}})} \sum_{k=1}^{J-n} \frac{(J-n-1)!}{k!(k-1)!(J-n-k)!} \frac{I_{J+k}(2\theta_{\text{rf}})}{\theta_{\text{rf}}^{J-S-1}} & \text{for } n < J \\ \frac{J!}{I_1(2\theta_{\text{rf}})} \frac{I_J(2\theta_{\text{rf}})}{\theta_{\text{rf}}^{J-1}} & \text{for } n = J \end{cases}$	(40)
$\mathbb{P}_{\text{pm}}^{\text{sm}}(S_n \theta_{\text{pm}}, J)$	$\binom{J}{n} \theta_{\text{pm}} \frac{(n-1)!(\theta_{\text{pm}})_{J_{\text{M}}-n}}{(\theta_{\text{pm}})_{J_{\text{M}}}}$	(41)
Local community	(Dispersal limitation)	
$\mathbb{P}_{\text{rf}}^{\text{sm}}(\vec{S} I, \theta_{\text{rf}}, J)$	$\binom{J}{\vec{n}} \frac{1}{(I)_J \prod_k S_k!} \sum_{a_1=1}^{n_1} \cdots \sum_{a_S=1}^{n_S} \left[ \prod_{i=1}^S \bar{s}(n_i, a_i) a_i! \right] I^A \theta_{\text{rf}}^{S-A} \frac{I_{A+S-1}(2\theta_{\text{rf}})}{I_1(2\theta_{\text{rf}})}$	(42)
$\mathbb{P}_{\text{pm}}^{\text{sm}}(\vec{S} I, \theta_{\text{pm}}, J)$	$\binom{J}{\vec{n}} \frac{1}{(I)_J \prod_k S_k!} \sum_{a_1=1}^{n_1} \cdots \sum_{a_S=1}^{n_S} \left[ \prod_{i=1}^S \bar{s}(n_i, a_i) (a_i - 1)! \right] I^A \frac{\theta_{\text{pm}}^S}{(\theta_{\text{pm}})_A}$	(43)
$\mathbb{P}_{\text{rf}}^{\text{sm}}(S_n I, \theta_{\text{rf}}, J)$	$\begin{cases} \binom{J}{n} \frac{1}{(I)_J} \sum_{k=1}^{J-n} \sum_{a_1=1}^n \sum_{a_2=k}^{J-n} \bar{s}(n, a_1) \bar{s}(J-n, a_2) \frac{a_1! a_2! (a_2-1)!}{k!(k-1)!(a_2-k)!} \frac{I^{a_1+a_2}}{\theta_{\text{rf}}^{a_1+a_2-k-1}} \frac{I_{a_1+a_2+k}(2\theta_{\text{rf}})}{I_1(2\theta_{\text{rf}})} & \text{for } n < J \\ \frac{1}{(I)_J} \sum_{a=1}^J \bar{s}(J, a) a! \frac{I^a}{\theta_{\text{rf}}^{a-1}} \frac{I_a(2\theta_{\text{rf}})}{I_1(2\theta_{\text{rf}})} & \text{for } n = J \end{cases}$	(44)
$\mathbb{P}_{\text{pm}}^{\text{sm}}(S_n I, \theta_{\text{pm}}, J)$	$\begin{cases} \binom{J}{n} \frac{\theta_{\text{pm}}}{(I)_J} \sum_{a_1=1}^n \sum_{a_2=1}^{J-n} \bar{s}(n, a_1) \bar{s}(J-n, a_2) (a_1-1)! \frac{I^{a_1+a_2}}{(\theta_{\text{pm}}+a_2)_{a_1}} & \text{for } n < J \\ \frac{\theta_{\text{pm}}}{(I)_J} \sum_{a=1}^n \bar{s}(J, a) (a-1)! \frac{I^a}{(\theta_{\text{pm}})_a} & \text{for } n = J \end{cases}$	(45)

For the expectation values we find (see “Appendix 6”):

$$\mathbb{E}^{\text{sm}}(S_n | I, \theta_{\text{rf}}, J) = \begin{cases} \binom{J}{n} \frac{1}{(I)_J} \sum_{k=1}^{J-n} \sum_{a_1=1}^n \sum_{a_2=k}^{J-n} \bar{s}(n, a_1) \bar{s}(J-n, a_2) & \text{for } n < J \\ \times \frac{a_1! a_2! (a_2 - 1)!}{k! (k-1)! (a_2 - k)!} \\ \times \frac{I^{a_1+a_2}}{\theta_{\text{rf}}^{a_1+a_2-k-1}} \frac{I_{a_1+a_2+k}(2\theta_{\text{rf}})}{I_1(2\theta_{\text{rf}})} & \\ \frac{1}{(I)_J} \sum_{a=1}^J \bar{s}(J, a) a! \frac{I^a}{\theta^{a-1}} \frac{I_a(2\theta_{\text{rf}})}{I_1(2\theta_{\text{rf}})} & \text{for } n = J \end{cases} \quad (46)$$

which is Eq. 44 in Table 2. This expression uses the approximation to  $\mathbb{E}^{\text{meta}}(S_n)$ . Using the exact formula would result in a much more complicated formula that is only negligibly different from the approximation (Fig. 1).

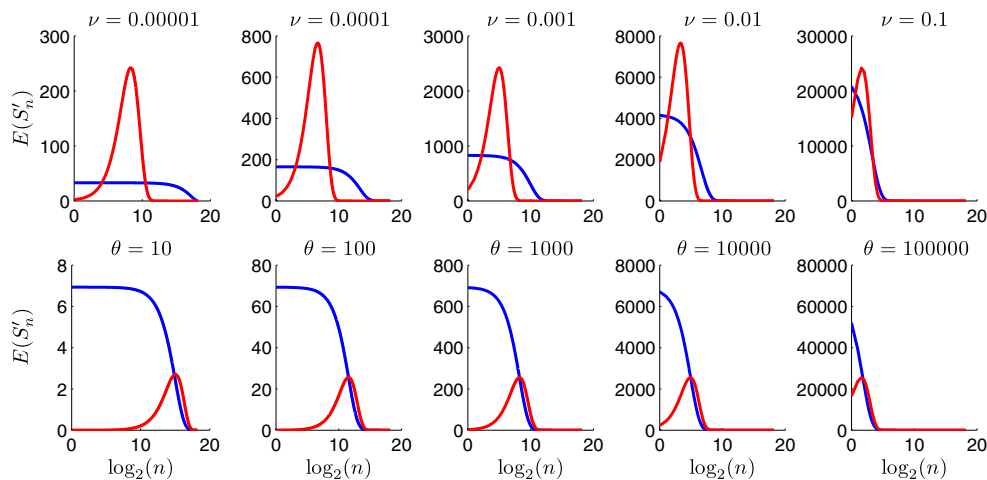
When dispersal is not limited (i.e.,  $I \rightarrow \infty$ ), then we have the random fission counterpart of the Ewens sampling formula (Ewens 1972). This formula is provided as Eq. 38 in Table 2. The associated expressions for the expected number of species with abundance  $n$  in the dispersal-unlimited sample is given in Table 2, as Eq. 40. Derivations can be found in “Appendices 4 and 5”. For comparison, Table 2 also shows the results for the point mutation model (Vallade and Houchmandzadeh 2003; Etienne 2005; Etienne and

Alonso 2005) which has  $\theta_{\text{pm}} = \nu_{\text{pm}} J_M$  where  $\nu_{\text{pm}}$  is the point speciation rate that is comparable to  $\nu_{\text{rf}}$ . Note that  $\theta_{\text{pm}}$  is often written as  $\frac{\tilde{\nu}_{\text{pm}}}{1 - \tilde{\nu}_{\text{pm}}} (J_M - 1)$  in the literature. In “Appendix 7”, we explain this difference, but here, we note that  $\nu_{\text{pm}}$  and  $\tilde{\nu}_{\text{pm}}$  are practically identical because they are very small and  $J_M$  is very large.

## Results

Random fission produces fewer abundant and fewer rare species, but more intermediately abundant species than point mutation with the same speciation rate  $\nu$  (Fig. 3). When  $\theta_{\text{rf}}$  is set equal to  $\theta_{\text{pm}}$ , then random fission produces fewer abundant species than point mutation (Fig. 3). With our definition of  $\theta_{\text{rf}}$ , the SAD of a sample becomes independent of metacommunity size, only being dependent on the compound parameter  $\theta_{\text{rf}}$ . This is similar to the role of  $\theta_{\text{pm}}$  in the point mutation model. There is another similarity between the two  $\theta$ s as well: when increasing  $\theta_{\text{pm}}$  and  $\theta_{\text{rf}}$ , the abundance curves for point mutation and random fission shift in the same way to higher abundances (Fig. 3).

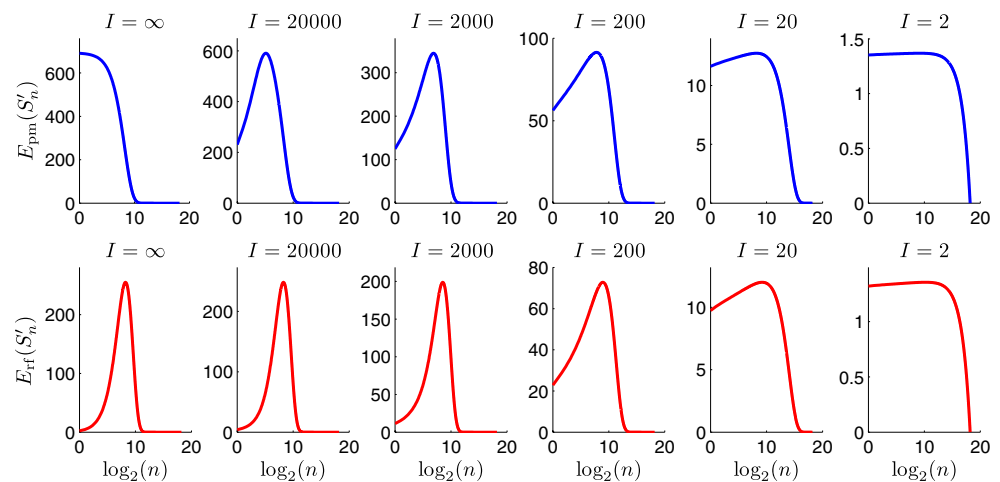
The total species richness in the metacommunity does not depend on  $J_M$  in large metacommunities with random fission speciation (Fig. 2). In contrast, in a metacommunity with point mutation speciation, metacommunity size always controls total species richness. Figure 4 shows the behavior of the SAD in a metacommunity with random fission, as a function of recruitment limitation. When recruitment is severely limited,



**Fig. 3** Comparison of the metacommunity SADs for random fission (Eq. 23, red) and point mutation (Eq. 24, blue) for various values of the speciation rate  $\nu$  (top row) and fundamental biodiversity number  $\theta$  (bottom row). Metacommunity size is set at

$J_M = 300,000$ . The SADs are defined as the expected number of species per unit of logarithmic abundance with base 2. In formula, this means that we plot  $\mathbb{E}^{\text{meta}}(S'_n) = n \ln 2 \times \mathbb{E}^{\text{meta}}(S_n)$  on the y-axis and the  $\log_2$  of abundance  $n$  on the x-axis

**Fig. 4** Comparison of the local community SADs (defined as in Fig. 3) for random fission (Eq. 44, red) and point mutation (Eq. 45, blue) for various values of recruitment limitation  $I$ . The fundamental biodiversity constant is set at  $\theta_{\text{pm}} = \theta_{\text{rf}} = 1,000$ . Sample size is  $J = 300,000$



the SADs of the point mutation and random fission models are very similar, suggesting that the mode of speciation does not leave a noticeable signature on the SAD in recruitment limited communities, but it does when recruitment limitation does not play a role.

As an illustration of our new sampling formula, we applied it to six tropical tree data sets (Volkov et al. 2005). Table 3 shows the parameter estimates obtained by likelihood maximization. It also shows the results obtained previously (Chave et al. 2006; Etienne et al. 2007b) for the point mutation model and compares the two models based on Akaike weights. Figure 5 shows the abundance distributions with the fitted models. Clearly the model with random fission speciation never performs significantly better than the point mutation model. There are two cases where the performance is similar.

First, Korup has as maximum likelihood estimates an infinite  $\theta$  and a very low  $m$ -value that corresponds to the Ewens estimate for  $I$ . This is the value of  $I$  that maximizes the Ewens sampling formula with parameter

$I$  (instead of  $\theta$ ); this value is well known to solve the equation:

$$S = \sum_{i=1}^J \frac{I}{I + i - 1} = I(\Psi(I + J) - \Psi(I)) \quad (47)$$

where  $\Psi$  is the psi or digamma function (see also Eq. 32 for a similar expression). For Korup, which has  $S = 308$  and  $J = 24591$ , this amounts to  $I = 49.5$ . When  $\theta$  is infinite, the mode of speciation is no longer important: Every immigrant in the local community will be of a new species. This is indistinguishable from a meta-community (ruled by point mutation) with  $\theta_{\text{pm}} = 49.5$  (Etienne et al. 2006).

Second, for Sinharaja, the point mutation and random fission models perform equally well, remarkably with almost identical  $m$ -values (but different  $\theta$ -values). This suggests more strongly that there is extreme recruitment limitation and metacommunity diversity is high, in contrast to the values reported by Volkov

**Table 3** Parameter estimates, maximum likelihood and model comparison for the fit of the two neutral models with different speciation modes to six local community abundance data sets

Site	$J^a$	$S^b$	Point mutation (pm) <sup>c</sup>			Random fission (rf) <sup>c</sup>			Comparison <sup>c</sup>	
			$\theta_{\text{pm}}$	$m$	ML <sup>d</sup>	$\theta_{\text{rf}}$	$m$	ML <sup>d</sup>	$w_{\text{pm}}$	$w_{\text{rf}}$
BCI	21,457	225	47.67	0.093	−308.73	595.1	0.0029	−311.92	0.96	0.04
Korup	24,591	308	52.73	0.547	−317.04	$\infty$	0.0020	−318.67	0.84	0.16
Pasoh	26,554	678	190.9	0.093	−359.38	1,528	0.0098	−363.75	0.99	0.01
Sinharaja	16,936	167	436.8	0.0019	−252.93	927.6	0.0019	−252.88	0.49	0.51
Yasuni	17,546	821	204.2	0.429	−297.15	10,980	0.0111	−306.75	1.00	0.00
Lambir	33,175	1,004	285.6	0.115	−386.38	2,500	0.0111	−402.32	1.00	0.00

<sup>a</sup>Sample size

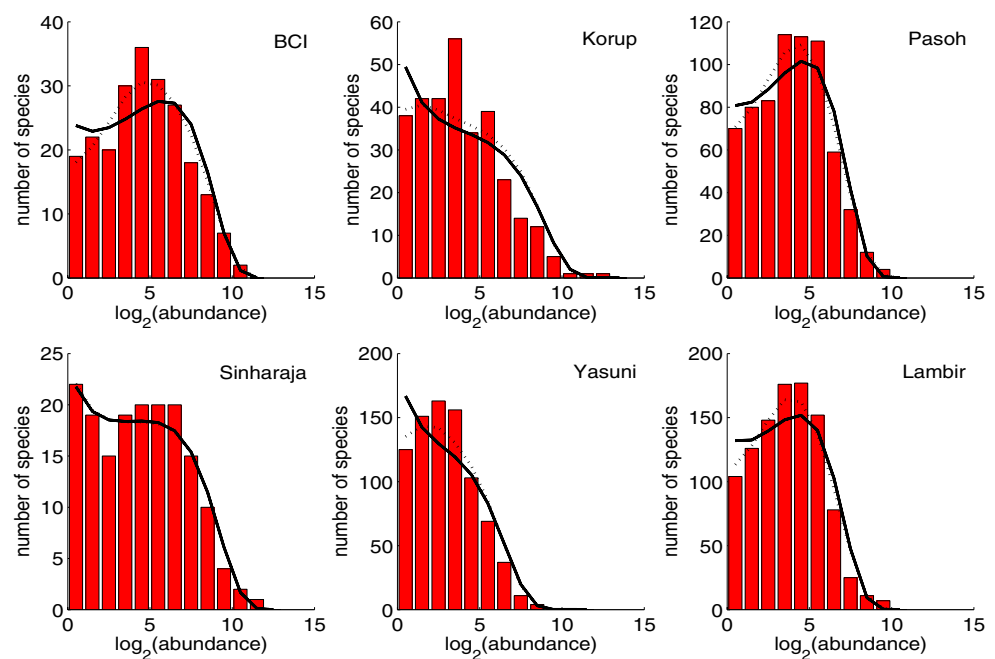
<sup>b</sup>Number of species observed

<sup>c</sup>These are the dispersal limited versions of the pm and rf models

<sup>d</sup>The logarithm of the maximum likelihood

<sup>e</sup>Comparison between point mutation and random fission in terms of Akaike weights  $w_i$

**Fig. 5** Species abundance distributions of the six tropical forest plots (bars) with the maximum likelihood expectations according to the point mutation (dotted line) and random fission (solid line) models. Maximum likelihoods and corresponding parameters are provided in Table 1. Binning is as in Fig. 1



et al. (2005), who missed the slightly higher likelihood optimum for high  $\theta$  and low  $m$  (Etienne et al. 2007b).

The random fission model does not seem to possess the dual optima exhibited by the point mutation model (Etienne et al. 2006). This is probably because the random fission sampling formula does not have the symmetry of the point mutation sampling formula where infinite  $\theta$  and infinite  $I$  have mathematically identical effects. Figure 6 shows the likelihood surface of the random fission model for BCI.

Although  $\theta_{rf}$  and  $\theta_{pm}$  are not directly comparable because they are defined differently, their ratio tells us

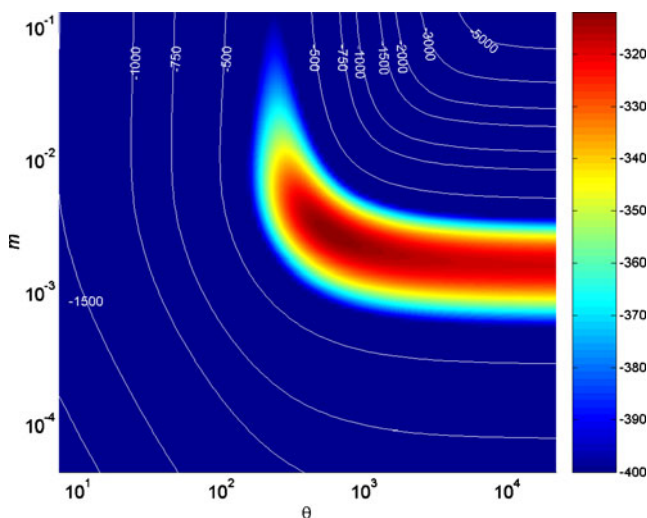
something about the relative values of the speciation rates in the two models:

$$\nu_{rf} < \nu_{pm} \text{ if } \nu_{pm} < \left( \frac{\theta_{pm}}{\theta_{rf}} \right)^2 \quad (48)$$

For Yasuni, the ratio of  $\theta_{pm}$  and  $\theta_{rf}$  is the smallest of all data sets (approximately  $\frac{1}{54}$ ) which means that  $\nu_{rf} < \nu_{pm}$  if  $\nu_{pm} < 0.0003$ . As values of  $\nu_{pm}$  larger than 0.0003 are highly unlikely, we can conclude that random fission must occur at a lower rate than point mutation to fit the observed SAD.

As a numerical example, assume that metacommunity size in the neotropics is of order  $J_M \approx 10^{10}$  (Ricklefs 2003; Nee 2005). This means that, for Yasuni, the corresponding speciation rates are  $\nu_{pm} \approx 10^{-8}$  and  $\nu_{rf} \approx 10^{-12}$ . The average species longevities are then (Ricklefs 2003)  $t_{pm} \approx -2 \ln(2\nu_{pm}) \approx 10^2$  and  $t_{rf} \approx \nu_{rf}^{-\frac{1}{2}} \approx 10^6$ , and the expected total richness is  $\mathbb{E}_{rf}^{meta}(S_M) \approx 10^4$  (see Eq. 31) and  $\mathbb{E}_{pm}^{meta}(S_M) \approx 4 \cdot 10^3$  (see Eq. 32). We discuss these results below. Note that all these speciation rates are community speciation rates divided by community death–birth events (remember we assumed the latter to be  $\mu = 1$ ). For an estimate of the species-level speciation rates measured in inverse generations, these numbers need to be multiplied by the average population size  $\frac{J_M}{\mathbb{E}_{meta}(S_M)}$ .

We also fitted the metacommunity sampling formula to two data sets of metacommunity abundances. The first consists of 41 plots of 1 ha in Panama (Condit et al. 2002), and the second consists of 50 plots of



**Fig. 6** Loglikelihood surface of the  $(\theta_{rf}, m)$ -parameter combination for the BCI data set

**Table 4** Parameter estimates, maximum likelihood, and model comparison for the fit of the two neutral models with different speciation modes to two metacommunity abundance data sets

Metacommunity	$J^a$	$S^b$	Point mutation (pm) <sup>c</sup>		Random fission (rf) <sup>c</sup>		Comparison <sup>c</sup>	
			$\theta_{pm}$	ML <sup>d</sup>	$\theta_{rf}$	ML <sup>d</sup>	$w_{pm}$	$w_{rf}$
Panama (pooled sample)	16,292	759	164.7	−311.09	795.8	−795.51	1.00	0.00
Panama (repeated samples)	41	35.4	119.8	−3.4644	252.3	−3.5348	0.48	0.52
Western Ghats (pooled sample)	19,555	304	50.99	−298.65	308.5	−524.03	1.00	0.00
Western Ghats (repeated samples)	50	30.4	51.45	−5.9220	116.3	−6.1143	0.53	0.47

*pooled sample* all individuals from all plots are put into one large sample, *repeated samples* many samples are taken with one individual from each plot and results are averaged over these samples

<sup>a</sup>Sample size

<sup>b</sup>Number of species observed; for the repeated samples this is the mean number of species

<sup>c</sup>These are the dispersal *unlimited* versions of the pm and rf models

<sup>d</sup>The logarithm of the maximum likelihood; for the repeated samples, this is the mean value

<sup>e</sup>Comparison between point mutation and random fission in terms of (mean) Akaike weights  $w_i$

1 ha in the Western Ghats in India (Munoz et al. 2007; Etienne 2009a). We did the analysis in two ways. First, we simply pooled the data of all the plots into a single sample. Second, we sampled a single individual from each plot, recorded the number of species, repeated this many times, and estimated  $\theta$  from the averaged number of species obtained (Etienne 2009a); loglikelihoods and Akaike weights were obtained by averaging over their values for each sample. The results are summarized in Table 4. In the first, “pooled sample”, analysis, the random fission model performed much worse than the point mutation model for both data sets. Yet, in the second, “repeated samples”, analysis, we find that the point mutation model performs only marginally better. We discuss this discrepancy below.

## Discussion

In this paper, we have presented a full sampling formula for the SAD in the neutral model with random fission speciation, with and without recruitment limitation. In contrast to Hubbell’s (2001) conjecture, the metacommunity abundance distribution (i.e., without recruitment limitation) is not a zero-sum multinomial. In fact, we have shown that the random fission mode of speciation produces a SAD that is identical to the broken-stick model of MacArthur (1957), when conditioned on total metacommunity size and total metacommunity species richness. An expression for the SAD (i.e.,  $\mathbb{P}^{\text{meta}}(\tilde{S}|J_M, S_M)$ ) for the discrete broken-stick (DBS) model was given in Etienne and Olff (2005), and this expression is mathematically identical to our Eq. 14. Our Eq. 18c is the continuous form of the broken-stick model (CBS), that is, in the limit of  $J_M \rightarrow \infty$ . Thus, random fission speciation provides a mechanistic explanation of this classic broken-stick model. Indeed, random fission and random stick-breaking are

equivalent processes mathematically, but is not immediately obvious that they lead to the same distribution because the random fission model includes ecological drift (random birth and death processes) whereas the broken-stick model is not clearly linked to biological processes.

Similarly, our dispersal-limited sampling formula can be interpreted as a dispersal-limited broken-stick model. Although the fits to data are not better than the point mutation model (which produces the logseries), adding dispersal limitation makes the fits much better than found by Etienne and Olff (2005) for the pure DBS model. We might therefore say that the broken-stick model has been resurrected to some extent from the natural death, so longed for by its inventor (MacArthur 1966). Cohen (1968) also showed that the broken stick could be produced by alternative models, but none of these were dynamical such as our birth–death–speciation model. The random fission model is also more general than these models because it also predicts an SAD without conditioning on total species number.

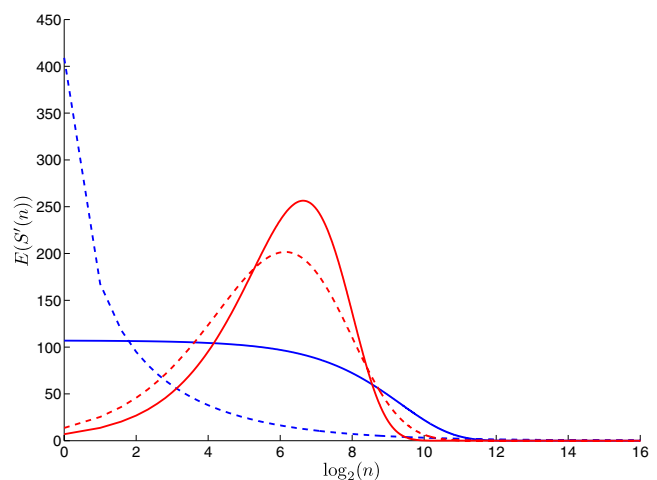
In contrast to another conjecture by Hubbell (2001), the metacommunity abundance distribution is not governed by two parameters ( $\nu_{rf}$  and  $J_M$ ), but by only one just like in the point mutation model. This parameter, the random fission fundamental biodiversity number, is defined as  $\theta_{rf} = \sqrt{\nu_{rf}} J_M$ . This definition of  $\theta_{rf}$  is different than that of  $\theta_{pm}$  by a factor of  $\sqrt{\nu_{rf}}$ . This explains why Hubbell (2001) thought that two independent parameters govern the random fission model: He found with his simulations that fixing  $\nu_{rf} J_M$  did not fully determine the abundance distribution. Had he fixed  $\sqrt{\nu_{rf}} J_M$ , then he would have found that the SAD was completely specified. Strictly, this is only true in the limit of large  $J_M$  that we are considering, but in practice, this limit is always approximating the system extremely well.



We have compared the fit of the random fission model against the point mutation model in three different ways. First, we confronted six local community data sets with the dispersal-limited sampling formulas. We found that the random fission model never performed significantly better than the point mutation model and did much worse in several cases. However, this may be due to nonneutral recruitment limitation (Jabot et al. 2008) obscuring the comparison of speciation modes. We therefore also compared fits of the models to metacommunity data directly. Unfortunately, large metacommunity data sets, where individuals are randomly sampled across a large spatial scale, are scarce. At best, there are a limited number of local samples, e.g., 41 and 50 in the data sets we used here. Although we recognize that the sampling effort for these data sets is appreciable, these data sets require some restrictive assumptions in their analysis. Our second comparison consisted of pooling all the samples as if the individuals were randomly sampled rather than sampled in local clusters, thereby ignoring spatial structure. Here, the random fission model performed miserably. Our third comparison consisted of repeatedly sampling one individual from each plot and averaging results over all these small samples (of sizes 41 and 50), thereby ignoring that the data showed an actual metacommunity richness that was much higher than that of each of those samples. In this case, random fission and point mutation performed equally well. Remarkably, the maximum likelihood estimates for  $\theta_{pm}$  are roughly of the same order of magnitude for the pooled sample and repeated samples, whereas the estimates for  $\theta_{rf}$  are three-fold smaller in the repeated samples than in the pooled samples. This suggests that the random fission model does not produce the correct scaling with sample size and even more so because the sample size in the pooled sample is larger than a true metacommunity sample with the same number of species, due to recruitment limitation. Thus, we conclude that the point mutation model really seems a much better description of the metacommunity than random fission model. The relatively reasonable fit of the random fission model to some of the local community data sets is only achieved by allowing for strong (and arguably unrealistic) dispersal limitation which eradicates the signature of the speciation mode (recall Fig. 4).

Fitting a model to a single snapshot of species abundances is a weak test of the model (McGill 2003; McGill et al. 2006, 2007) because many different mechanisms may lead to similar patterns in SADs (Cohen 1968; Purves and Pacala 2005). However, it is a test that a serious model should pass. Not all models produce a realistic SAD and failure to do so is a strong reason

to reject a model. Here, we have confirmed earlier findings (Etienne et al. 2007b) that SADs can have resolving power with respect to explanatory models, and particularly the speciation mode leaves a clear signature on the SAD (see Fig. 7, which will be discussed further below). We have shown that the random fission mode never performs significantly better than the point mutation mode; indeed, it performs much worse for almost all data sets and the metacommunity data sets in particular. This suggests that point mutation is a more likely mode of speciation in tropical forests than random fission. However, random fission predicts more reasonable speciation rates, metacommunity richness, and species longevity based on the fitted parameters. Perhaps the estimates for species richness seem somewhat high, suggesting that the metacommunity for the Yasuni plot is of the order of the entire Amazon, but such a continental extent of the metacommunity may be quite reasonable taking into account recent evidence for long-distance dispersal (Jabot and Chave 2009). Thus, neither of the two models is capable of making more accurate predictions of various quantities or patterns. For now, this leads us to reject both models in their current, spatially implicit, form as plausible explanations of community structure. We do want to



**Fig. 7** Metacommunity SADs (as defined in Fig. 3) for four different speciation modes: individual-level point mutation (Eq. 24, solid blue line), species-level point mutation (Etienne et al. 2007b, dashed blue line), individual-level random fission (Eq. 8/Eq. 23, solid red line), and species-level random fission (dashed red line). The solution for  $\mathbb{E}^{\text{meta}}(S_n)$  in the latter case is, like the individual-level random fission model, given by Eq. 8 but with  $s_i = \frac{v_i}{J_M}$ . Metacommunity size is set at  $J_M = 100,000$  and expected metacommunity richness is set at  $\mathbb{E}^{\text{meta}}(S_M) = 1,000$ . Expected metacommunity richness is the area under the curve for all models because of our definition of the SAD:  $\int \mathbb{E}^{\text{meta}}(S'_n) d \log_2 n = \int \mathbb{E}^{\text{meta}}(S_n) n \ln 2 d \log_2 n = \int \mathbb{E}^{\text{meta}}(S_n) dn = \mathbb{E}^{\text{meta}}(S_M)$



caution against too rapid definite conclusions for the following reasons.

The performance of a model in producing realistic SADs depends on all ingredients of the model (Etienne 2007), not just the one under consideration, such as the speciation mode in this paper. We have studied the speciation mode in a neutral context, so the performance is also affected by the neutrality assumption. Random fission or point mutation or both may perform better or worse in a nonneutral setting, or even have negligible influence. Zillio and Condit (2007) have, however, found that many communities, neutral or non-neutral, are primarily driven by the process of how new species enter the community. In our opinion, this justifies our study of the influence of speciation in the simplest context where species differences do not play a role, but we do not rule out the possibility that the mode of speciation may lead to more or less realistic predictions under nonneutral conditions. Only by formulating different models and comparing their predictions to data can we get a feel for the behavior of these models and for how informative SADs are for inferring mechanism.

Speciation is not the only way new species can enter the community: Immigration is another. Interestingly, the sampling formula resulting from point mutation speciation also describes long-distance dispersal (see, e.g., Etienne et al. 2007a) which may leave a more profound signature than speciation. Only a model that includes both long-distance dispersal and speciation will be able to distinguish the relative contribution of these processes. An analytically tractable spatially implicit model with long-distance dispersal and random fission speciation is not yet available, let alone a spatially explicit model. Nevertheless, a spatially explicit neutral model with long-distance (fat-tailed) dispersal and point mutation speciation *has* been studied (Rosindell and Cornell 2009) using coalescence techniques (Rosindell et al. 2008). It has been shown that speciation rates required to yield predictions on species–area relationships that are in agreement with observations are much lower (i.e., more realistic) than in models with Gaussian dispersal kernels (Rosindell and Cornell 2007). This suggests that long-distance dispersal is an important force shaping ecological communities and can mimic the effect of speciation, but does not rule out that speciation still leaves a signature on community structure. Unfortunately, coalescence does not seem compatible with random fission speciation because random fission is not time reversible, and therefore, this powerful simulation method cannot be

employed to further study the effect of random fission speciation on community structure in a spatially explicit context. Perhaps new analytical techniques based on functional differential equations (O'Dwyer et al. 2009) will prove useful.

Random fission assigns a very low probability to samples with a few very abundant species, even more so than point mutation. This explains why the fit to data is worse for the former speciation mode and the fact that the best fitted model does not seem to follow unimodal SAD data in some cases (even though the unimodal shape is more characteristic for random fission than for point mutation, see Figs. 3 and 4). Etienne et al. (2007b) pointed out that a visual goodness-of-fit estimate may be deceiving and logtransformed the data and predictions to make this clear. Here, logtransformation supports the same conclusion (not shown). At the same time, random fission predicts fewer rare species in the metacommunity than point mutation and therefore produces a more lognormal-like shaped SAD which is believed to be a better description by many ecologists (Preston 1948, 1962; McGill 2003). The fit to the metacommunity data sets suggests that this advantage does not offset the disadvantage at large abundances. In fact, this advantage may not be so advantageous after all because the random fission model even seems to predict too few rare species.

For Hubbell's neutral models  $\nu$  is a constant speciation rate per individual (for both point mutation and random fission), in contrast to common practice in speciation research where the speciation rate is usually a rate per species (Stanley 1979; Etienne and Apol 2009). The results in this paper apply to Hubbell's model. Assuming a constant speciation rate per species in the neutral model with point mutation has also been studied (Etienne et al. 2007b), and it was found that this assumption can make a large difference. Therefore, the remaining combination of random fission speciation with a constant speciation rate at the species level appears to be a necessary last step to get a complete picture. In fact, while a constant speciation rate per individual seems a logical first choice for the individual-level process of point mutation, a constant speciation rate per species seems more obvious for the species-level process of random fission. The community will be less diverse than in the individual-level case because abundant species will be less likely to undergo fission. For the same reason, the species-level point mutation model produced less diversity than the individual-level point mutation model (Etienne et al. 2007a). Figure 7 shows a numerical comparison of all four speciation

modes. One observes that species-level random fission is similar to individual-level random fission but has a bit more abundant and rare species. Because very abundant species are not only more likely than in individual-level random fission model but also than in the individual-level point mutation model (this can only be seen after logtransforming Fig. 7), it might be that a model with species-level random fission speciation provides a good fit to SAD data as well as reasonable predictions for speciation rates and species longevities. We have not been able to find a sampling formula for this model, so this remains an open question.

We have presented analytical results for the spatially implicit model with the random fission mode of speciation. These are now on a par with analytical results for point mutation, allowing future studies of community models to compare both alternatives.

**Acknowledgements** We thank Franck Jabot and a few anonymous reviewers for their constructive comments. RSE thanks the financial support of The Netherlands Organization for Scientific Research (NWO).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## Appendix 1: Derivation of the dynamics of $\mathbb{E}^{\text{meta}}(S_n)$ and the corresponding stationary distribution

The dynamics of  $\mathbb{E}^{\text{meta}}(S_n)$  can be split into two parts

$$\frac{d}{dt}\mathbb{E}^{\text{meta}}(S_n) = \frac{d^{\text{DB}}}{dt}\mathbb{E}^{\text{meta}}(S_n) + \frac{d^{\text{S}}}{dt}\mathbb{E}^{\text{meta}}(S_n) \quad (49)$$

First, we consider the first part: dynamics due to death–birth events. The transition rate is given by

$$R_{k,i}^{\text{DB}}(\vec{S}) = \begin{cases} \mu \frac{kS_k}{J_M} \frac{iS_i}{J_M - 1} & \text{if } k \neq i \\ \mu \frac{k^2 S_k (S_k - 1)}{J_M (J_M - 1)} & \text{if } k = i. \end{cases} \quad (50)$$

The case  $k = i$  takes into account only those events in which the dead and newborn individual belong to a different species. If not, the death–birth event has no net effect.

Species with abundance  $n$  disappear ( $S'_n - S_n = -1$ ) in transitions  $R_{k,i}^{\text{DB}}$

with  $k = n$ : an individual in a species with abundance  $n$  dies  
with  $i = n$ : an individual in a species with abundance  $n$  reproduces

Species with abundance  $n$  appear ( $S'_n - S_n = 1$ ) in transitions  $R_{k,i}^{\text{DB}}$

with  $k = n + 1$ : an individual in a species with abundance  $n + 1$  dies,  
with  $i = n - 1$ : an individual in a species with abundance  $n - 1$  reproduces.

Hence,

$$\begin{aligned} \frac{d^{\text{DB}}}{dt}\mathbb{E}^{\text{meta}}(S_n) &= \sum_{\vec{S}} \sum_{k,i} \mathbb{P}^{\text{meta}}(\vec{S}) R_{k,i}^{\text{DB}}(\vec{S}) \\ &\quad \times ((\delta_{k,n} + \delta_{i,n})(-1) + (\delta_{k,n+1} + \delta_{i,n-1})(+1)) \\ &= \sum_{\vec{S}} \mathbb{P}^{\text{meta}}(\vec{S}) \left( \sum_k \left( R_{k,n-1}^{\text{DB}}(\vec{S}) - R_{k,n}^{\text{DB}}(\vec{S}) \right) \right. \\ &\quad \left. + \sum_{\ell} \left( R_{n+1,\ell}^{\text{DB}}(\vec{S}) - R_{n,\ell}^{\text{DB}}(\vec{S}) \right) \right) \\ &= \sum_{\vec{S}} \mathbb{P}^{\text{meta}}(\vec{S}) (r_{n-1}S_{n-1} + r_{n+1}S_{n+1} - 2r_nS_n) \\ &= r_{n-1}\mathbb{E}^{\text{meta}}(S_{n-1}) + r_{n+1}\mathbb{E}^{\text{meta}}(S_{n+1}) \\ &\quad - 2r_n\mathbb{E}^{\text{meta}}(S_n) \end{aligned} \quad (51)$$

where we defined

$$r_n = \mu \frac{n(J_M - n)}{J_M(J_M - 1)} \quad (52)$$

Note that we implicitly assumed that  $n > 1$  and  $n < J_M$ ; the extension to the cases  $n = 1$  and  $n = J_M$  is straightforward.

Next, we consider the second part of Eq. 49: dynamics due to speciation events. The transition rate is given by

$$R_{k,i}^{\text{S}}(\vec{S}) = \nu \frac{kS_k}{J_M} \frac{1}{k-1}. \quad (53)$$

Species with abundance  $n$  disappear ( $S'_n - S_n = -1$ ) in transitions  $R_{k,i}^{\text{S}}(\vec{S})$  with  $k = n$ . Species with abundance  $n$  appear ( $S'_n - S_n = 1$ ) in transitions  $R_{k,i}^{\text{S}}(\vec{S})$  with  $n = i$  or  $n = k - i$ . Note that for  $k > n$  and  $k \neq 2n$ , there are two fission events that lead to a new species with abundance  $n$  and that for  $k = 2n$ , there is one fission

event that leads to two new species with abundance  $n$ . Hence,

$$\begin{aligned} \frac{d^S}{dt} \mathbb{E}^{\text{meta}}(S_n) &= \sum_{\vec{S}} \sum_{k,i} \mathbb{P}^{\text{meta}} \\ &\quad \times \left( \vec{S} \right) R_{k,i}^S \left( \vec{S} \right) (\delta_{k,n}(-1) + (\delta_{i,n} + \delta_{i,k-n})(+1)) \\ &= \sum_{\vec{S}} \mathbb{P}^{\text{meta}} \left( \vec{S} \right) \\ &\quad \times \left( \sum_i -R_{n,i}^S \left( \vec{S} \right) + \sum_k \left( R_{k,n}^S \left( \vec{S} \right) + R_{k,k-n}^S \left( \vec{S} \right) \right) \right) \\ &= \sum_{\vec{S}} \mathbb{P}^{\text{meta}} \left( \vec{S} \right) \left( -s_n S_n + \sum_{k>n} \frac{2}{k-1} s_k S_k \right) \\ &= -s_n \mathbb{E}^{\text{meta}}(S_n) + \sum_{k>n} \frac{2}{k-1} s_k \mathbb{E}^{\text{meta}}(S_k) \end{aligned} \quad (54)$$

where we defined

$$s_n = v \frac{n}{J_M} \quad (55)$$

Note that we implicitly assumed that  $n > 1$  and  $n < J_M$ ; the extension to the cases  $n = 1$  and  $n = J_M$  is straightforward.

Summing Eqs. 51 and 54 yields

$$\begin{aligned} \frac{d}{dt} \mathbb{E}^{\text{meta}}(S_1) &= -2r_1 \mathbb{E}^{\text{meta}}(S_1) + r_2 \mathbb{E}^{\text{meta}}(S_2) \\ &\quad + \sum_{i=2}^{J_M} \frac{2}{i-1} s_i \mathbb{E}^{\text{meta}}(S_i) \end{aligned} \quad (56a)$$

$$\begin{aligned} \frac{d}{dt} \mathbb{E}^{\text{meta}}(S_n) &= r_{n-1} \mathbb{E}^{\text{meta}}(S_{n-1}) - (2r_n + s_n) \mathbb{E}^{\text{meta}}(S_n) \\ &\quad + r_{n+1} \mathbb{E}^{\text{meta}}(S_{n+1}) + \sum_{i=n+1}^{J_M} \frac{2}{i-1} s_i \mathbb{E}^{\text{meta}}(S_i) \\ &\quad \text{for } 1 < n < J_M \end{aligned} \quad (56b)$$

$$\begin{aligned} \frac{d}{dt} \mathbb{E}^{\text{meta}}(S_{J_M}) &= r_{J_M-1} \mathbb{E}^{\text{meta}}(S_{J_M-1}) \\ &\quad - (r_{J_M} + s_{J_M}) \mathbb{E}^{\text{meta}}(S_{J_M}) \end{aligned} \quad (56c)$$

Equations 56a, 56b, and 56c keep the total number of individuals in the community constant, as required:

$$\frac{d}{dt} \sum_{n=1}^{J_M} n \mathbb{E}^{\text{meta}}(S_n) = 0 \quad (57)$$

Furthermore, the total expected species richness satisfies

$$\begin{aligned} \frac{d}{dt} \mathbb{E}^{\text{meta}}(S) &= \frac{d}{dt} \sum_{n=1}^{J_M} \mathbb{E}^{\text{meta}}(S_n) \\ &= -r_1 \mathbb{E}^{\text{meta}}(S_1) + \sum_{n=2}^{J_M} s_n \mathbb{E}^{\text{meta}}(S_n) \end{aligned} \quad (58)$$

where the first term describes species extinction and the second term describes speciation.

Equations 56a, 56b, and 56c are closed in  $\mathbb{E}^{\text{meta}}(S_n)$ , that is, they do not contain expectations such as  $\mathbb{E}^{\text{meta}}(S_n S_m)$ . This seems to be a general property of neutral models (Vanpeteghem et al. 2008). Similarly, we can derive the dynamical equations for the second-order moments  $\mathbb{E}^{\text{meta}}(S_n S_m)$ , which are again closed, that is, they do not contain third- or higher-order moments.

The equilibrium solution of Eqs. 56a, 56b, and 56c can be found most easily by setting Eqs. 56b, 56c, and 58 to zero, remembering that  $\sum_{n=1}^{J_M} n \mathbb{E}^{\text{meta}}(S_n) = J_M$  and then solving for  $a_i$  defined by

$$a_i = \frac{\mathbb{E}^{\text{meta}}(S_n)}{\mathbb{E}^{\text{meta}}(S_{n-1})} \quad (59)$$

This yields the following expression for  $\mathbb{E}^{\text{meta}}(S_n)$ :

$$\mathbb{E}^{\text{meta}}(S_n) = \prod_{i=1}^n a_i \quad (60a)$$

with

$$\begin{cases} a_1 = \frac{v}{r_1 + s_1} \\ a_i = \frac{r_{i-1}}{2r_i + s_i - r_{i+1}a_{i+1} - \sum_{j=i+1}^{J_M} s_j \frac{2}{j-1} \prod_{k=i+1}^j a_k} \\ \text{for } 1 < i < J_M \\ a_{J_M} = \frac{r_{J_M-1}}{r_{J_M} + s_{J_M}} \end{cases} \quad (60b)$$

where  $a_1$  results from setting Eq. 58 to zero,  $a_{J_M}$  from setting Eq. 56c to zero, and  $a_i$  for  $1 < i < J_M$  by setting Eq. 56b to zero.

## Appendix 2: Derivation of $\mathbb{E}^{\text{meta}}(S_n|J_M)$ from the Ansatz

$\mathbb{E}^{\text{meta}}(S_n|J_M)$  follows in a straightforward way from the Ansatz for  $\mathbb{P}^{\text{meta}}(\vec{S}|J_M)$ :

$$\begin{aligned}\mathbb{E}^{\text{meta}}(S_n|J_M) &= \sum_{\vec{S}|J_M} S_n \mathbb{P}^{\text{meta}}(\vec{S}|J_M) \\ &= \frac{1}{Z(vJ_M, J_M)} \sum_{\vec{S}|J_M} S_n \prod_k \frac{(vJ_M)^{S_k}}{S_k!} \\ &= \frac{vJ_M}{Z(vJ_M, J_M)} \sum_{\vec{S}|J_M} \frac{(vJ_M)^{S_n-1}}{(S_n-1)!} \prod_{k \neq n} \frac{(vJ_M)^{S_k}}{S_k!} \\ &= \frac{vJ_M}{Z(vJ_M, J_M)} \sum_{\vec{S}|J_M-n} \prod_k \frac{(vJ_M)^{S_k}}{S_k!} \\ &= vJ_M \frac{Z(vJ_M, J_M - n)}{Z(vJ_M, J_M)}\end{aligned}\quad (61)$$

## Appendix 3: Derivation of the distribution, mean, and variance of the total number of species $S_M$ for large $\theta_{\text{rf}}$

We first compute the mean and variance of the number of species  $S_M$  in the metacommunity for large  $\theta_{\text{rf}}$ . We have

$$\begin{aligned}\mathbb{E}^{\text{meta}}(S_M|J_M \rightarrow \infty) &= \frac{1}{Z(\theta_{\text{rf}})} \sum_{S_M=1}^{\infty} \frac{S_M \theta_{\text{rf}}^{2S_M}}{S_M! (S_M-1)!} \\ &= \theta_{\text{rf}} \frac{I_0(2\theta_{\text{rf}})}{I_1(2\theta_{\text{rf}})}\end{aligned}\quad (62a)$$

$$\begin{aligned}\mathbb{E}^{\text{meta}}(S_M^2|J_M \rightarrow \infty) &= \frac{1}{Z(\theta_{\text{rf}})} \sum_{S_M=1}^{\infty} \frac{S_M^2 \theta_{\text{rf}}^{2S_M}}{S_M! (S_M-1)!} \\ &= \theta_{\text{rf}}^2 + \theta_{\text{rf}} \frac{I_0(2\theta_{\text{rf}})}{I_1(2\theta_{\text{rf}})}\end{aligned}\quad (62b)$$

The asymptotic behavior of the Bessel functions  $I_\alpha(x)$  is given by

$$\begin{aligned}I_\alpha(x) &= \frac{e^x}{\sqrt{2\pi x}} \left( 1 - \frac{4\alpha^2 - 1}{8} x^{-1} + \mathcal{O}(x^{-2}) \right) \\ &\text{as } x \rightarrow \infty.\end{aligned}\quad (63)$$

where we use the big  $\mathcal{O}$  notation to describe the asymptotic behavior of a function. Hence,

$$\frac{I_0(x)}{I_1(x)} = 1 + \frac{1}{2} x^{-1} + \mathcal{O}(x^{-2}) \quad \text{as } x \rightarrow \infty \quad (64)$$

so that

$$\begin{aligned}\mathbb{E}^{\text{meta}}(S_M|J_M \rightarrow \infty) &= \theta_{\text{rf}} + \frac{\theta_{\text{rf}}^0}{4} + \mathcal{O}(\theta_{\text{rf}}^{-1}) \\ &\text{as } \theta_{\text{rf}} \rightarrow \infty\end{aligned}\quad (65a)$$

$$\begin{aligned}\mathbb{E}^{\text{meta}}(S_M^2|J_M \rightarrow \infty) &= \theta_{\text{rf}}^2 + \theta_{\text{rf}} + \mathcal{O}(\theta_{\text{rf}}^0) \\ &\text{as } \theta_{\text{rf}} \rightarrow \infty.\end{aligned}\quad (65b)$$

For the variance,

$$\begin{aligned}\text{Var}^{\text{meta}}(S_M|J_M \rightarrow \infty) &= \mathbb{E}^{\text{meta}}(S_M^2|J_M \rightarrow \infty) - (\mathbb{E}^{\text{meta}}(S_M|J_M \rightarrow \infty))^2 \\ &= \theta_{\text{rf}}^2 + \theta_{\text{rf}} - \theta_{\text{rf}}^2 - \frac{\theta_{\text{rf}}}{2} + \mathcal{O}(\theta_{\text{rf}}^0) \\ &= \frac{\theta_{\text{rf}}}{2} + \mathcal{O}(\theta_{\text{rf}}^0) \quad \text{as } \theta_{\text{rf}} \rightarrow \infty.\end{aligned}\quad (66)$$

The asymptotic behavior of the distribution  $\mathbb{P}^{\text{meta}}(S_M|J_M \rightarrow \infty)$  for the number of species  $S_M$  can be analyzed further. Taking the logarithm of Eq. 15a,

$$\begin{aligned}\ln(\mathbb{P}^{\text{meta}}(S_M|J_M \rightarrow \infty)) &= (2S_M - 1) \ln \theta_{\text{rf}} - \ln I_1(2\theta_{\text{rf}}) - \ln S_M! - \ln(S_M - 1)! \\ &\quad (67)\end{aligned}$$

and using asymptotic formulas for the Bessel function and the factorial (Stirling's formula),

$$\ln I_\alpha(x) = x + \mathcal{O}(\ln x) \quad \text{as } x \rightarrow \infty \quad (68a)$$

$$\ln(n!) = n \ln n - n + \mathcal{O}(\ln n) \quad \text{as } n \rightarrow \infty, \quad (68b)$$

we get

$$\begin{aligned}\ln(\mathbb{P}^{\text{meta}}(S_M|J_M \rightarrow \infty)) &= 2S_M \ln \theta_{\text{rf}} - 2\theta_{\text{rf}} - 2S_M \ln S_M \\ &\quad + 2S_M + \mathcal{O}(\ln S_M, \ln \theta_{\text{rf}}) \\ &\quad \text{as } S_M, \theta_{\text{rf}} \rightarrow \infty.\end{aligned}\quad (69)$$

The expression in the first line, considered as a function of  $S_M$ , has a maximum at  $\theta_{\text{rf}}$ . Developing this function up to second order in  $S_M - \theta_{\text{rf}}$ , we get

$$\begin{aligned}\ln(\mathbb{P}^{\text{meta}}(S_M|J_M \rightarrow \infty)) &= -\frac{(S_M - \theta_{\text{rf}})^2}{\theta_{\text{rf}}} \\ &\quad + \mathcal{O}((S_M - \theta_{\text{rf}})^3)\end{aligned}\quad (70)$$

and hence,

$$\mathbb{P}^{\text{meta}}(S_M|J_M \rightarrow \infty) \sim e^{-\frac{(S_M - \theta_{\text{rf}})^2}{\theta_{\text{rf}}}} \quad \text{for large } \theta_{\text{rf}}. \quad (71)$$

Thus, for large  $\theta_{\text{rf}}$  the distribution  $\mathbb{P}^{\text{meta}}(S_M|J_M \rightarrow \infty)$  for the number of species  $S_M$  is normally distributed, with mean  $\theta_{\text{rf}}$  and standard deviation  $\sqrt{\frac{\theta_{\text{rf}}}{2}}$ .

#### Appendix 4: Derivation of the approximation to $\mathbb{E}^{\text{meta}}(S_n|J_M)$ and $\mathbb{E}^{\text{meta}}(S_k S_\ell|J_M)$

We start by computing the expectations of the number of species with abundance  $n$  in the metacommunity and of  $S_k S_\ell$  conditional on metacommunity size  $J_M$  and total number of species  $S_M$ . First, consider the trivial case that  $S_M = 1$ . The abundance of the only species present is equal to the metacommunity size. Hence,

$$\begin{aligned}\mathbb{E}^{\text{meta}}(S_n|J_M, S_M = 1) &= 0 \quad \text{for } n < J_M \\ \mathbb{E}^{\text{meta}}(S_{J_M}|J_M, S_M = 1) &= 1\end{aligned}\quad (72)$$

Next, consider the case that  $S_M \geq 2$ . We have

$$\begin{aligned}\mathbb{E}^{\text{meta}}(S_n|J_M, S_M) &= \sum_{\vec{S}|J_M, S_M} S_n \mathbb{P}^{\text{meta}}(\vec{S}|J_M, S_M) \\ &= \frac{S_M!}{(J_M-1)!} \sum_{\vec{S}|J_M, S_M} \frac{S_n}{\prod_k S_k!} \\ &= \frac{S_M}{(J_M-1)!} \sum_{\vec{S}|J_M, S_M} \frac{(S_M-1)!}{(S_n-1)! \prod_{k \neq n} S_k!} \\ &= \frac{S_M}{(J_M-1)!} \binom{J_M-n-1}{S_M-2} \\ &\quad \times \sum_{\vec{S}|J_M-n, S_M-1} \mathbb{P}^{\text{meta}}(\vec{S}|J_M-n, S_M-1) \\ &= \frac{S_M}{(J_M-1)!} \binom{J_M-n-1}{S_M-2}.\end{aligned}\quad (73)$$

Assuming that  $J_M$  is large we find

$$\begin{aligned}\mathbb{E}^{\text{meta}}(S_n|J_M, S_M, J_M \gg S_M) &= S_M \frac{(S_M-1)!(J_M-S_M)!}{(J_M-1)!} \\ &\quad \times \frac{(J_M-n-1)!}{(S_M-2)!(J_M-n-S_M+1)!} \\ &\approx S_M(S_M-1) \frac{(J_M-n)^{S_M-2}}{J_M^{S_M-1}} \\ &= \frac{S_M(S_M-1)}{J_M} \left(1 - \frac{n}{J_M}\right)^{S_M-2}\end{aligned}\quad (74)$$

Similarly, one can compute the expectation of higher-order moments of  $S_n$ . We illustrate this for the expectation  $\mathbb{E}^{\text{meta}}(S_k S_\ell|J_M, S_M)$  with  $k \neq \ell$ . First, consider the case with  $S_M = 2$ . The abundances of the two

species present should sum up to the metacommunity size,  $k + \ell = J_M$ . Hence,

$$\begin{aligned}\mathbb{E}^{\text{meta}}(S_k S_\ell|J_M, S_M = 2) &= 0 \quad \text{if } k + \ell \neq J_M \\ \mathbb{E}^{\text{meta}}(S_k S_\ell|J_M, S_M = 2) &= \frac{2}{J_M-1} \quad \text{if } k + \ell = J_M.\end{aligned}\quad (75)$$

Next, consider the case with  $S_M \geq 3$ . We have

$$\begin{aligned}\mathbb{E}^{\text{meta}}(S_k S_\ell|J_M, S_M) &= \sum_{\vec{S}|J_M, S_M} S_k S_\ell \mathbb{P}^{\text{meta}}(\vec{S}|J_M, S_M) \\ &= \frac{S_M!}{(J_M-1)!} \sum_{\vec{S}|J_M, S_M} \frac{S_k S_\ell}{\prod_m S_m!} \\ &= \frac{S_M(S_M-1)}{(J_M-1)!} \sum_{\vec{S}|J_M, S_M} \frac{(S_M-2)!}{(S_k-1)!(S_\ell-1)! \prod_{m \neq k, \ell} S_m!} \\ &= \frac{S_M(S_M-1)}{(J_M-1)!} \binom{J_M-k-\ell-1}{S_M-3} \\ &\quad \times \sum_{\vec{S}|J_M-k-\ell, S_M-2} \mathbb{P}^{\text{meta}}(\vec{S}|J_M-k-\ell, S_M-2) \\ &= \frac{S_M(S_M-1)}{(J_M-1)!} \binom{J_M-k-\ell-1}{S_M-3}.\end{aligned}\quad (76)$$

Assuming that  $J_M$  is large leads to

$$\begin{aligned}\mathbb{E}^{\text{meta}}(S_k S_\ell|J_M, S_M, J_M \gg S_M) &= S_M(S_M-1) \frac{(S_M-1)!(J_M-S_M)!}{(J_M-1)!} \\ &\quad \times \frac{(J_M-k-\ell-1)!}{(S_M-3)!(J_M-k-\ell-S_M+2)!} \\ &\approx S_M(S_M-1)^2(S_M-2) \frac{(J_M-k-\ell)^{S_M-3}}{J_M^{S_M-1}} \\ &= \frac{S_M(S_M-1)^2(S_M-2)}{J_M^2} \left(1 - \frac{k+\ell}{J_M}\right)^{S_M-3}\end{aligned}\quad (77)$$

Now we are ready to compute the same quantities without conditioning on  $S_M$ . The number of species with  $J_M$  individuals,  $S_{J_M}$ , is different from zero only if there is a single species in the community. The expected number of species  $\mathbb{E}(S_{J_M}|J_M)$  is therefore equal to the probability that there is only one species in the community,  $P(S_M = 1) = \frac{\theta_{\text{rt}}}{J_1(2\theta_{\text{rt}})}$ . To compute the expectation  $\mathbb{E}(S_n|J_M)$  for  $n < J_M$ , we only have to consider com-

munities with at least two species. Assuming that  $J_M$  is large, we have

$$\begin{aligned}
 \mathbb{E}(S_n|J_M) &\approx \sum_{S_M=2}^{J_M} \mathbb{E}(S_n|J_M, S_M) \mathbb{P}(S_M|J_M \rightarrow \infty) \\
 &\approx \sum_{S_M=2}^{\infty} \frac{1}{J_M} \left(1 - \frac{n}{J_M}\right)^{S_M-2} \frac{1}{Z(\theta_{\text{rf}})} \frac{\theta_{\text{rf}}^{2S_M}}{(S_M-1)!(S_M-2)!} \\
 &= \sum_{S_M=1}^{\infty} \frac{1}{J_M} \left(1 - \frac{n}{J_M}\right)^{S_M-1} \frac{1}{Z(\theta_{\text{rf}})} \frac{\theta_{\text{rf}}^{2S_M+2}}{S_M!(S_M-1)!} \\
 &= \frac{1}{J_M} \frac{\theta_{\text{rf}}^2}{Z(\theta_{\text{rf}})} \left(1 - \frac{n}{J_M}\right)^{-1} \sum_{S_M=1}^{\infty} \frac{\left(\left(1 - \frac{n}{J_M}\right) \theta_{\text{rf}}^2\right)^{S_M}}{S_M!(S_M-1)!} \\
 &= \frac{\theta_{\text{rf}}^2}{J_M \sqrt{1 - \frac{n}{J_M}}} \frac{I_1\left(2\theta_{\text{rf}} \sqrt{1 - \frac{n}{J_M}}\right)}{I_1(2\theta_{\text{rf}})} \quad (78)
 \end{aligned}$$

The product  $S_k S_\ell$  with  $k + \ell = J_M$  is different from zero only for communities with two species, one with abundance  $k$  and the other with abundance  $\ell$ . Hence, the expectation  $\mathbb{E}(S_k S_\ell | J_M)$  with  $k + \ell = J_M$  has only contributions from two-species communities. To compute the expectation  $\mathbb{E}(S_k S_\ell | J_M)$  with  $k + \ell < J_M$ , we only have to consider communities with at least three species. Assuming that  $J_M$  is large, we have

$$\begin{aligned}
 \mathbb{E}(S_k S_\ell | J_M) &\approx \sum_{S_M=3}^{J_M} \mathbb{E}(S_k S_\ell | J_M, S_M) \mathbb{P}(S_M|J_M \rightarrow \infty) \\
 &\approx \sum_{S_M=3}^{\infty} \frac{1}{J_M^2} \left(1 - \frac{k+\ell}{J_M}\right)^{S_M-3} \frac{1}{Z(\theta_{\text{rf}})} \frac{\theta_{\text{rf}}^{2S_M}}{(S_M-2)!(S_M-3)!} \\
 &= \sum_{S_M=1}^{\infty} \frac{1}{J_M^2} \left(1 - \frac{k+\ell}{J_M}\right)^{S_M-1} \frac{1}{Z(\theta_{\text{rf}})} \frac{\theta_{\text{rf}}^{2S_M+4}}{S_M!(S_M-1)!} \\
 &= \frac{1}{J_M^2 \left(1 - \frac{k+\ell}{J_M}\right)} \frac{\theta_{\text{rf}}^4}{Z(\theta_{\text{rf}})} \sum_{S_M=1}^{\infty} \frac{\left(\left(1 - \frac{k+\ell}{J_M}\right) \theta_{\text{rf}}^2\right)^{S_M}}{S_M!(S_M-1)!} \\
 &= \frac{\theta_{\text{rf}}^4}{J_M^2 \sqrt{1 - \frac{k+\ell}{J_M}}} \frac{I_1\left(2\theta_{\text{rf}} \sqrt{1 - \frac{k+\ell}{J_M}}\right)}{I_1(2\theta_{\text{rf}})} \quad (79)
 \end{aligned}$$

## Appendix 5: Derivation of the sampling formula

$$\mathbb{P}^{\text{smp}}(\vec{S}|I, \theta_{\text{rf}}, J)$$

The derivation starts with the general formula for a dispersal-limited sample (Etienne and Alonso 2005) where we make the conditioning on  $S_M$  explicit:

$$\begin{aligned}
 \mathbb{P}^{\text{smp}}(\vec{S}|S_M, I, \theta_{\text{rf}}, J) &= \left(\frac{J}{n}\right) \frac{1}{(I)_J} \int_{\sum \vec{p}=1} \prod_{i=1}^{S_M} (Ip_i)_{n_i} \rho(\vec{p}|S_M) d\vec{p} \quad (80)
 \end{aligned}$$

By multiplying with  $\mathbb{P}^{\text{meta}}(S_M|J_M \rightarrow \infty)$  and summing over all  $S_M$ , we find

$$\begin{aligned}
 \mathbb{P}^{\text{smp}}(\vec{S}|I, \theta_{\text{rf}}, J) &= \sum_{S_M=S}^{\infty} \mathbb{P}^{\text{meta}}(S_M|J_M \rightarrow \infty) \frac{S_M!}{(S_M-S)! \prod_{k=1}^J S_k!} \\
 &\quad \times P(\vec{S}|S_M, I, \theta_{\text{rf}}, J) \\
 &= \left(\frac{J}{n}\right) \frac{1}{(I)_J} \sum_{S_M=S}^{\infty} \mathbb{P}^{\text{meta}}(S_M|J_M \rightarrow \infty) \\
 &\quad \times \frac{S_M!}{(S_M-S)! \prod_{k=1}^J S_k!} \\
 &\quad \times \int_{\sum \vec{p}=1} \prod_{i=1}^{S_M} (Ip_i)_{n_i} \rho^{\text{meta}}(\vec{p}|S_M) d\vec{p} \quad (81)
 \end{aligned}$$

We can now substitute Eq. 18c and use the Stirling number formulation of the Pochhammer symbol, Eq. 34,

$$\begin{aligned}
 \mathbb{P}^{\text{smp}}(\vec{S}|I, \theta_{\text{rf}}, J) &= \left(\frac{J}{n}\right) \frac{1}{(I)_J \prod_k S_k!} \sum_{S_M=S}^{\infty} \mathbb{P}^{\text{meta}}(S_M|J_M \rightarrow \infty) \\
 &\quad \times \frac{S_M!(S_M-1)!}{(S_M-S)!} \sum_{a_1=1}^{n_1} \dots \sum_{a_S=1}^{n_S} \left[ \prod_{i=1}^S \bar{s}(n_i, a_i) \right] I^A \\
 &\quad \times \int_{\sum \vec{p}=1} \left[ \prod_{i=1}^{S_M} p_i^{a_i} \right] d\vec{p} \quad (82)
 \end{aligned}$$

with  $A = \sum_i a_i$ . Next we evaluate the integral,

$$\begin{aligned}
 \int_{\sum \vec{p}=1} \left[ \prod_{i=1}^{S_M} p_i^{a_i} \right] d\vec{p} &= \frac{\prod_{i=1}^{S_M} a_i!}{(A + S_M - 1)!} \\
 &= \frac{\prod_{i=1}^S a_i!}{(A + S_M - 1)!} \quad (83)
 \end{aligned}$$



because  $a_i = 0$  for all species that are not present in the sample. Substituting this and Eq. 12, we obtain

$$\begin{aligned} \mathbb{P}^{\text{smp}}(\vec{S}|I, \theta_{\text{rf}}, J) &= \binom{J}{\vec{n}} \frac{1}{(I)_J \prod_k S_k!} \sum_{S_M=S}^{\infty} \frac{\theta_{\text{rf}}^{2S_M}}{Z(\theta_{\text{rf}})(S_M - S)!} \\ &\quad \times \sum_{a_1=1}^{n_1} \dots \sum_{a_S=1}^{n_S} \left[ \prod_{i=1}^S \bar{s}(n_i, a_i) \right] \frac{I^A \prod_{i=1}^S a_i!}{(A + S_M - 1)!} \end{aligned} \quad (84)$$

The sum over  $S_M$  can be expressed in terms of the modified Bessel function of the first kind,

$$\sum_{S_M=S}^{\infty} \frac{\theta_{\text{rf}}^{2S_M}}{(S_M - S)!(A + S_M - 1)!} = \theta_{\text{rf}}^{S-A+1} I_{A+S-1}(2\theta_{\text{rf}}) \quad (85)$$

Substituting this and Eq. 15b in Eq. 84, we obtain our final result (Eq. 33):

$$\begin{aligned} \mathbb{P}^{\text{smp}}(\vec{S}|I, \theta_{\text{rf}}, J) &= \binom{J}{\vec{n}} \frac{1}{(I)_J \prod_k S_k!} \sum_{a_1=1}^{n_1} \dots \sum_{a_S=1}^{n_S} \left[ \prod_{i=1}^S \bar{s}(n_i, a_i) a_i! \right] \\ &\quad \times I^A \theta_{\text{rf}}^{S-A} \frac{I_{A+S-1}(2\theta_{\text{rf}})}{I_1(2\theta_{\text{rf}})} \end{aligned} \quad (86)$$

This formula can be evaluated numerically, similarly to the sampling formula for the point mutation model (Etienne 2005). If there is no dispersal limitation (i.e., let  $I \rightarrow \infty$ ), we have

$$\begin{aligned} \mathbb{P}^{\text{smp}}(\vec{S}|I \rightarrow \infty, \theta_{\text{rf}}, J) &= \binom{J}{\vec{n}} \sum_{S_M=S}^{\infty} \mathbb{P}^{\text{meta}}(S_M|J_M \rightarrow \infty) \frac{S_M!}{(S_M - S)! \prod_{k=1}^J S_k!} \\ &\quad \times \int_{\sum \vec{p}=1} \prod_{i=1}^{S_M} p_i^{n_i} \rho^{\text{meta}}(\vec{p}|S_M) d\vec{p} \\ &= \binom{J}{\vec{n}} \sum_{S_M=S}^{\infty} \frac{\theta_{\text{rf}}^{2S_M}}{S_M! (S_M - 1)!} \\ &\quad \times \frac{S_M! (S_M - 1)!}{(S_M - S)! \prod_{k=1}^J S_k!} \frac{\prod_{i=1}^S n_i!}{(J + S_M - 1)!} \\ &= \frac{J!}{\prod_{k=1}^J S_k!} \frac{1}{\theta_{\text{rf}} I_1(2\theta_{\text{rf}})} \sum_{S_M=S}^{\infty} \frac{\theta_{\text{rf}}^{2S_M}}{(S_M - S)! (J + S_M - 1)!} \\ &= \frac{J!}{\prod_{k=1}^J S_k!} \theta_{\text{rf}}^{S-J} \frac{I_{J+S-1}(2\theta_{\text{rf}})}{I_1(2\theta_{\text{rf}})} \end{aligned} \quad (87)$$

where in the third line we have used Eq. 85.

## Appendix 6: Derivation of $\mathbb{E}^{\text{smp}}(S_n|I, \theta_{\text{rf}}, J)$

We derive  $\mathbb{E}^{\text{smp}}(S_n|I, \theta_{\text{rf}}, J)$  (for  $n < J$ ) by following Alonso and McKane (2004) and Etienne and Alonso (2005). We take a dispersal-limited sample from the metacommunity where the density of species with relative abundance  $p$  in the metacommunity with  $S_M$  species is given by  $S_M \rho^{\text{meta}}(p|S_M)$ :

$$\begin{aligned} \mathbb{E}^{\text{smp}}(S_{n < J}|I, \theta_{\text{rf}}, J) &= \binom{J}{n} \sum_{S_M=1}^{\infty} \mathbb{P}^{\text{meta}}(S_M|J_M \rightarrow \infty) \\ &\quad \times \int_0^1 \frac{(Ip)_n (I(1-p))_{J-n}}{(I)_J} S_M \rho^{\text{meta}}(p|S_M) dp \\ &= \binom{J}{n} \sum_{S_M=1}^{\infty} \frac{1}{Z(\theta_{\text{rf}})} \frac{\theta_{\text{rf}}^{2S_M}}{S_M! (S_M - 1)!} \\ &\quad \times \int_0^1 \frac{(Ip)_n (I(1-p))_{J-n}}{(I)_J} S_M (S_M - 1) (1-p)^{S_M-2} dp \\ &= \binom{J}{n} \frac{1}{(I)_J Z(\theta_{\text{rf}})} \sum_{a_1=1}^n \sum_{a_2=1}^{J-n} \bar{s}(n, a_1) \bar{s}(J-n, a_2) I^{a_1+a_2} \\ &\quad \times \sum_{S_M=2}^{\infty} \frac{\theta_{\text{rf}}^{2S_M}}{(S_M - 1)! (S_M - 2)!} \int_0^1 p^{a_1} (1-p)^{a_2+S_M-2} dp \\ &= \binom{J}{n} \frac{1}{(I)_J Z(\theta_{\text{rf}})} \sum_{a_1=1}^n \sum_{a_2=1}^{J-n} \bar{s}(n, a_1) \bar{s}(J-n, a_2) \Gamma(a_1 + 1) \\ &\quad \times I^{a_1+a_2} \sum_{S_M=2}^{\infty} \frac{\theta_{\text{rf}}^{2S_M}}{(S_M - 1)! (S_M - 2)!} \frac{\Gamma(a_2 + S_M - 1)}{\Gamma(a_1 + a_2 + S_M)} \\ &= \binom{J}{n} \frac{\theta_{\text{rf}}^4}{(I)_J Z(\theta_{\text{rf}})} \sum_{a_1=1}^n \sum_{a_2=1}^{J-n} \bar{s}(n, a_1) \bar{s}(J-n, a_2) \\ &\quad \times \frac{a_1! a_2!}{(a_1 + a_2 + 1)!} \\ &\quad \times I_1^{a_1+a_2} {}_1F_2(a_2 + 1, \{2, a_1 + a_2 + 2\}, \theta_{\text{rf}}^2) \end{aligned} \quad (88)$$

We can also retain the integral (instead of writing it in terms of Stirling numbers) and write

$$\begin{aligned} \mathbb{E}^{\text{smp}}(S_{n < J}|I, \theta_{\text{rf}}, J) &= \binom{J}{n} \frac{1}{Z(\theta_{\text{rf}})} \int_0^1 \frac{(Ip)_n (I(1-p))_{J-n}}{(I)_J} (1-p)^{-2} \\ &\quad \times \sum_{S_M=2}^{\infty} \frac{(\theta_{\text{rf}}^2 (1-p))^{S_M}}{(S_M - 1)! (S_M - 2)!} dp \end{aligned}$$

$$= \binom{J}{n} \frac{\theta_{\text{rf}}^3}{Z(\theta_{\text{rf}})} \int_0^1 \frac{(Ip)_n (I(1-p))_{J-n}}{(I)_J} \frac{1}{\sqrt{1-p}} \times I_1 \left( 2\theta_{\text{rf}} \sqrt{1-p} \right) dp \quad (89)$$

Alternatively, we can directly use our sampling formula to obtain

$$\begin{aligned} \mathbb{E}^{\text{smp}}(S_{n < J} | I, \theta_{\text{rf}}, J) &= \sum_{\vec{S}} \mathbb{P}^{\text{smp}}(\vec{S} | I, \theta_{\text{rf}}, J) S_n \\ &= \binom{J}{n} \frac{1}{(I)_J I_1(2\theta_{\text{rf}})} \sum_{k=1}^{J-n} \sum_{a_1=1}^n \sum_{a_2=k}^{J-n} \tilde{s}(n, a_1) \tilde{s}(J-n, a_2) \\ &\quad \times \frac{a_1! a_2! (a_2-1)!}{k! (k-1)! (a_2-k)!} I^{a_1+a_2} \frac{I_{a_1+a_2+k} (2\theta_{\text{rf}})}{\theta_{\text{rf}}^{a_1+a_2+k-1}} \quad (90) \end{aligned}$$

All these expressions can be shown to be mathematically identical.

If there is no dispersal limitation, these equations can be simplified. Equation 88 reduces to

$$\begin{aligned} \mathbb{E}^{\text{smp}}(S_n | I \rightarrow \infty, \theta_{\text{rf}}, J) &= \binom{J}{n} \sum_{S_M=1}^{\infty} \mathbb{P}^{\text{meta}}(S_M | J_M \rightarrow \infty) \\ &\quad \times \int_0^1 p^n (1-p)^{J-n} S_M \rho(p | S_M) dp \\ &= \binom{J}{n} \sum_{S_M=1}^{\infty} \frac{1}{Z(\theta_{\text{rf}})} \frac{\theta_{\text{rf}}^{2S_M}}{S_M! (S_M-1)!} \\ &\quad \times \int_0^1 p^n (1-p)^{J-n} S_M (S_M-1) (1-p)^{S_M-2} dp \\ &= \binom{J}{n} \frac{1}{Z(\theta_{\text{rf}})} \sum_{S_M=2}^{\infty} \frac{\theta_{\text{rf}}^{2S_M}}{(S_M-1)! (S_M-2)!} \\ &\quad \times \int_0^1 p^n (1-p)^{J-n+S_M-2} dp \\ &= \binom{J}{n} \frac{1}{Z(\theta_{\text{rf}})} \sum_{S_M=2}^{\infty} \frac{\theta_{\text{rf}}^{2S_M}}{(S_M-1)! (S_M-2)!} \\ &\quad \times \frac{\Gamma(J-n+S_M-1) \Gamma(n+1)}{\Gamma(J+S_M)} \\ &= \frac{\theta_{\text{rf}}^3}{J+1} \frac{{}_1F_2(J-n+1, \{2, J+2\}, \theta_{\text{rf}}^2)}{I_1(2\theta_{\text{rf}})} \quad (91) \end{aligned}$$

where  ${}_1F_2(a, \{b, c\}, x)$  is the generalized hypergeometric function.

Likewise, Eq. 89 reduces to

$$\mathbb{E}^{\text{smp}}(S_n | I \rightarrow \infty, \theta_{\text{rf}}, J) = \binom{J}{n} \frac{\theta_{\text{rf}}^3}{Z(\theta_{\text{rf}})} \int_0^1 \frac{p^n (1-p)^{J-n}}{\sqrt{1-p}} \times I_1 \left( 2\theta_{\text{rf}} \sqrt{1-p} \right) dp \quad (92)$$

and Eq. 90 reduces to

$$\begin{aligned} \mathbb{E}^{\text{smp}}(S_n | I \rightarrow \infty, \theta_{\text{rf}}, J) &= \frac{J!}{I_1(2\theta_{\text{rf}})} \sum_{k=1}^{J-n} \frac{(J-n-1)!}{k! (k-1)! (J-n-k)!} \frac{I_{J+k} (2\theta_{\text{rf}})}{\theta_{\text{rf}}^{J-k-1}} \quad (93) \end{aligned}$$

## Appendix 7: The fundamental biodiversity number $\theta$

We treated the death–birth process and the speciation process as decoupled processes. The point mutation mode of speciation, however, was originally envisioned as a birth event where a crucial mutation had caused a new species to arise, and therefore, speciation and birth are related to one another. If the death–birth rate is  $\tilde{\mu}$  and the probability of a birth resulting in an individual of a new species is  $\tilde{v}_{\text{pm}}$ , then the transition rates for the point mutation model can be written as:

$$R^{\text{DB}}(\vec{S}, \vec{S}_{k,i}^{\text{DB}}) = R_{k,i}^{\text{DB}}(\vec{S}) = \tilde{\mu} (1 - \tilde{v}_{\text{pm}}) \frac{k S_k}{J_M} \frac{i S_i}{J_M - 1} \quad (94)$$

and

$$R_{\text{pm}}^{\text{S}}(\vec{S}, \vec{S}_{k,i}^{\text{S}}) = R_{k,i}^{\text{S}}(\vec{S}) = \tilde{\mu} \tilde{v}_{\text{pm}} \frac{k S_k}{J_M} \delta_1(i) \quad (95)$$

These transition rates are identical to the ones in Eqs. 3 and 6 if we set

$$\mu = \tilde{\mu} (1 - \tilde{v}_{\text{pm}}) \quad (96a)$$

$$\nu_{\text{pm}} = \tilde{\mu} \tilde{v}_{\text{pm}} \quad (96b)$$

Hence, we observe that a simple transformation of the parameters makes the equations identical.

The fundamental biodiversity number  $\theta$  is defined in the point mutation model as the number of speciation events per birth event, or the ratio of the speciation rate and the per capita birth rate. Because the per capita birth rate is the overall death–birth rate divided by the total number of individuals after a death event,  $J_M - 1$ , we have for the model where the processes of speciation and death–birth are decoupled, the following formula for  $\theta_{\text{pm}}$

$$\theta_{\text{pm}} = \frac{\nu_{\text{pm}}}{\frac{\mu}{J_M-1}} = \frac{\nu_{\text{pm}} (J_M - 1)}{\mu} \approx \frac{\nu_{\text{pm}} J_M}{\mu} \quad (97)$$

while in the model where they are related,  $\theta_{\text{pm}}$  is expressed as

$$\theta_{\text{pm}} = \frac{\tilde{\mu}\tilde{\nu}_{\text{pm}}}{\frac{\tilde{\mu}(1-\tilde{\nu}_{\text{pm}})}{J_{\text{M}}-1}} = \frac{\tilde{\nu}_{\text{pm}}}{1-\tilde{\nu}_{\text{pm}}} (J_{\text{M}} - 1) \quad (98)$$

With Eq. 96, we see that these formulas coincide.

In the random fission model, we defined the fundamental biodiversity number in a different way: We used the appropriate scaling of  $\nu$  and  $J_{\text{M}}$  in the limit of  $J_{\text{M}} \rightarrow \infty$ ,  $\nu_{\text{rf}} \rightarrow 0$  which is in this case such that  $\theta_{\text{rf}} = \sqrt{\nu_{\text{rf}}} J_{\text{M}}$  is finite. In the point mutation model, we could also have defined  $\theta_{\text{pm}}$  by the appropriate scaling which is in this case such that  $\theta_{\text{pm}} = \nu_{\text{pm}} J_{\text{M}}$ . Hence, for point mutation, the scaling definition practically coincides with the definition as the number of speciation events per birth event, but for random fission, these definitions are different. We chose this scaling definition because with this definition the species abundance distribution behaves similarly (Fig. 3).

## References

- Allen AP, Savage VM (2007) Setting the absolute tempo of biodiversity dynamics. *Ecol Lett* 10:637–646
- Alonso D, McKane AJ (2004) Sampling Hubbell's neutral theory of biodiversity. *Ecol Lett* 7:901–910
- Belyea LR, Lancaster J (1999) Assembly rules within a contingent ecology. *Oikos* 86:402–416
- Caswell H (1976) Community structure: a neutral model analysis. *Ecol Monogr* 46:327–354
- Cavender-Bares J, Kozak K, Fine P, Kembel S (2009) The merging of community ecology and phylogenetic biology. *Ecol Lett* 12:693–715
- Chave J, Alonso D, Etienne RS (2006) Comparing models of species abundance. *Nature* 441:E1–E2
- Cody M, JM Diamond E (1975) Ecology and evolution of communities. Belknap, Cambridge
- Cohen JE (1968) Alternate derivations of a species-abundance relation. *Am Nat* 102:165–172
- Colwell RK, Winkler DW (1984) A null model for null models in biogeography. In: Strong LADR, Simberloff D, Thistle AB (eds) Ecological communities: conceptual issues and the evidence. Princeton University Press, Princeton
- Condit R, Pitman N, Leigh EG, Chave J, Terborgh J, Foster RB, Nunez P, Aguilar S, Valencia R, Villa G, Muller-Landau HC, Losos, E, Hubbell SP (2002) Beta-diversity in tropical forest trees. *Science* 295:666–669
- Conlisk J, Conlisk E, Harte J (2010) Hubbell's local abundance distribution: insights from a simple colonization rule. *Oikos* 119:379–383
- De Aguiar MAM, Baranger M, Baptestini EM, Kaufman L, Bar-Yam Y (2009) Global patterns of speciation and diversity. *Nature* 460:384–388
- Etienne RS (2005) A new sampling formula for neutral biodiversity. *Ecol Lett* 8:253–260
- Etienne RS (2007) A neutral sampling formula for multiple samples and an "exact" test of neutrality. *Ecol Lett* 10:608–618
- Etienne RS (2009a) Improved estimation of neutral model parameters for multiple samples with different degrees of dispersal limitation. *Ecology* 90:847–852
- Etienne RS (2009b) Maximum likelihood estimation of neutral model parameters for multiple samples with different degrees of dispersal limitation. *J Theor Biol* 257:510–514
- Etienne RS, Alonso D (2005) A dispersal-limited sampling theory for species and alleles. *Ecol Lett* 8:1147–1156
- Etienne RS, Alonso D (2007) Neutral community theory: how stochasticity and dispersal-limitation can explain species coexistence. *J Stat Phys* 128:485–510
- Etienne RS, Apol MEF (2009) Estimating speciation and extinction rates from diversity data and the fossil record. *Evolution* 63:244–255
- Etienne RS, Olff H (2005) Confronting different models of community structure to species-abundance data: a Bayesian model comparison. *Ecol Lett* 8:493–504
- Etienne RS, Latimer AM, Silander JA, Cowling RM (2006) Comment on "Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot". *Science* 311:610b
- Etienne RS, Alonso D, McKane AJ (2007a) The zero-sum assumption in neutral biodiversity theory. *J Theor Biol* 248:522–536
- Etienne RS, Apol MEF, Olff H, Weissing FJ (2007b) Modes of speciation and the neutral theory of biodiversity. *Oikos* 116:241–258
- Ewens WJ (1972) The sampling theory of selectively neutral alleles. *Theor Popul Biol* 3:87–112
- Haegeman B, Etienne RS (2008) Relaxing the zero-sum assumption in neutral biodiversity theory. *J Theor Biol* 252:288–294
- Haegeman B, Etienne RS (2009) Neutral models with generalised speciation. *Bull Math Biol* 71:1507–1519
- Haegeman B, Etienne RS (2010) A self-consistent approach to neutral community speciation models. *Phys Rev E* 81:031911
- Hubbell SP (2001) The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton
- Jabot F, Chave J (2009) Inferring the parameters of the neutral theory of biodiversity using phylogenetic information and implications for tropical forests. *Ecol Lett* 12:239–248
- Jabot F, Etienne RS, Chave J (2008) Reconciling neutral community models and environmental filtering: theory and an empirical test. *Oikos* 117:1308–1320
- MacArthur RH (1957) On the relative abundance of bird species. *Proc Natl Acad Sci U S A* 43:293–295
- MacArthur R (1966) Note on Mrs Pielou's comments. *Ecology* 47:1074
- MacArthur RH, Wilson EO (1967) Island biogeography. Princeton University Press, Princeton
- McGill BJ (2003) Strong and weak tests of macroecological theory. *Oikos* 102:679–685
- McGill BJ, Maurer BA, Weiser MD (2006) Empirical evaluation of neutral theory. *Ecology* 87:1411–1426
- McGill B, Etienne RS, Gray JS, Alonso D, Anderson MJ, Benecha HK, Dornelas M, Enquist BJ, Green JL, He F, Hurlbert A, Magurran A, Marquet P, Maurer B, Ostling A, Soykan CU, Ugland K, White EP (2007) Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol Lett* 10:995–1015
- Mouillot D, Gaston K (2007) Geographical range size heritability: what do neutral models with different modes of speciation predict? *Glob Ecol Biogeogr* 16:367–380
- Munoz F, Couteron P, Ramesh BR, Etienne RS (2007) Inferring parameters of neutral communities: from one single large to several small samples. *Ecology* 88:2482–2488

- Nee S (2005) The neutral theory of biodiversity: do the numbers add up? *Funct Ecol* 19:173–176
- O'Dwyer JP, Lake JK, Ostling A, Savage VM, Green JL (2009) An integrative framework for stochastic, size-structured community assembly. *Proc Natl Acad Sci* 106:6170–6175
- Preston FW (1948) The commonness, and rarity, of species. *Ecol* 29:254–283
- Preston FW (1962) The canonical distribution of commonness and rarity. Parts I and II. *Ecology* 43:185–215, 410–432
- Pueyo S (2006) Diversity: between neutrality and structure. *Oikos* 112:392–405
- Purves DW, Pacala SW (2005) Ecological drift in niche-structured communities: neutral pattern does not imply neutral process. In: Burslem D, Pinard M, Hartley S (eds) *Biotic interactions in the tropics*. Cambridge University Press, Cambridge, pp 107–138
- Ricklefs R (1987) Community diversity: relative roles of local and regional processes. *Science* 235:167–171
- Ricklefs RE (2003) A comment on Hubbell's zero-sum ecological drift model. *Oikos* 100:185–192
- Rosindell JL, Cornell SJ (2007) Species–area relationships from a spatially explicit neutral model in an infinite landscape. *Ecol Lett* 10:586–595
- Rosindell JL, Cornell SJ (2009) Species–area curves, neutral models and long distance dispersal. *Ecology* 90:1743–1750
- Rosindell J, Wong Y, Etienne RS (2008) A coalescence approach to spatial neutral ecology. *Ecol Inform* 3:259–271
- Rosindell J, Cornell SJ, Hubbell SP, Etienne RS (2010) Protracted speciation revitalizes the neutral theory of biodiversity. *Ecol Lett* (in press)
- Stanley SM (1979) *Macroevolution: pattern and process*. Freeman, San Francisco
- Vallade M, Houchmandzadeh B (2003) Analytical solution of a neutral model of biodiversity. *Phys Rev E* 68:061902
- Vanpeteghem D, Zemb O, Haegeman B (2008) Dynamics of neutral biodiversity. *Math Biosci* 212:88–98
- Volkov I, Banavar JR, Hubbell SP, Maritan A (2003) Neutral theory and relative species abundance in ecology. *Nature* 424:1035–1037
- Volkov I, Banavar JR, He F, Hubbell SP, Maritan A (2005) Density dependence explains tree species abundance and diversity in tropical forests. *Nature* 438:658–661
- Volkov I, Banavar JR, Hubbell SP, Maritan A (2007) Patterns of relative species abundance in rainforests and coral reefs. *Nature* 450:45–49
- Webb C, Ackerly D, McPeck M, Donoghue M (2002) Phylogenies and community ecology. *Ann Rev Ecol Syst* 33:475–505
- Weiher E, Keddy P (2001) *Ecological assembly rules: perspectives, advances, retreats*. Cambridge University Press, Cambridge
- Zillio T, Condit R (2007) The impact of neutrality, niche differentiation and species input on diversity and abundance distributions. *Oikos* 116:931–940